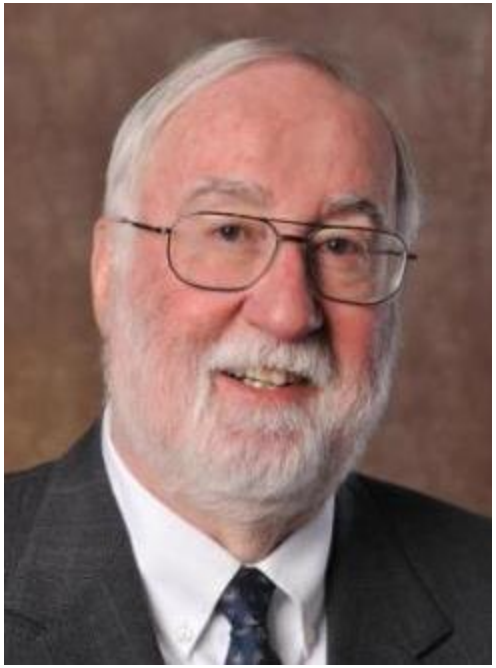
Public Health Statistics

Evidence Based Medicine

- *"Half of what you'll learn in medical school will be shown to be either dead wrong or out of date within five years of your graduation; the trouble is that nobody can tell you which half—so the most important thing to learn is how to learn on your own."*



David Sackett, MD

BMJ 12-18-03 http://www.bmj.com/content/327/7429/1430.full

Where is the Wisdom…?

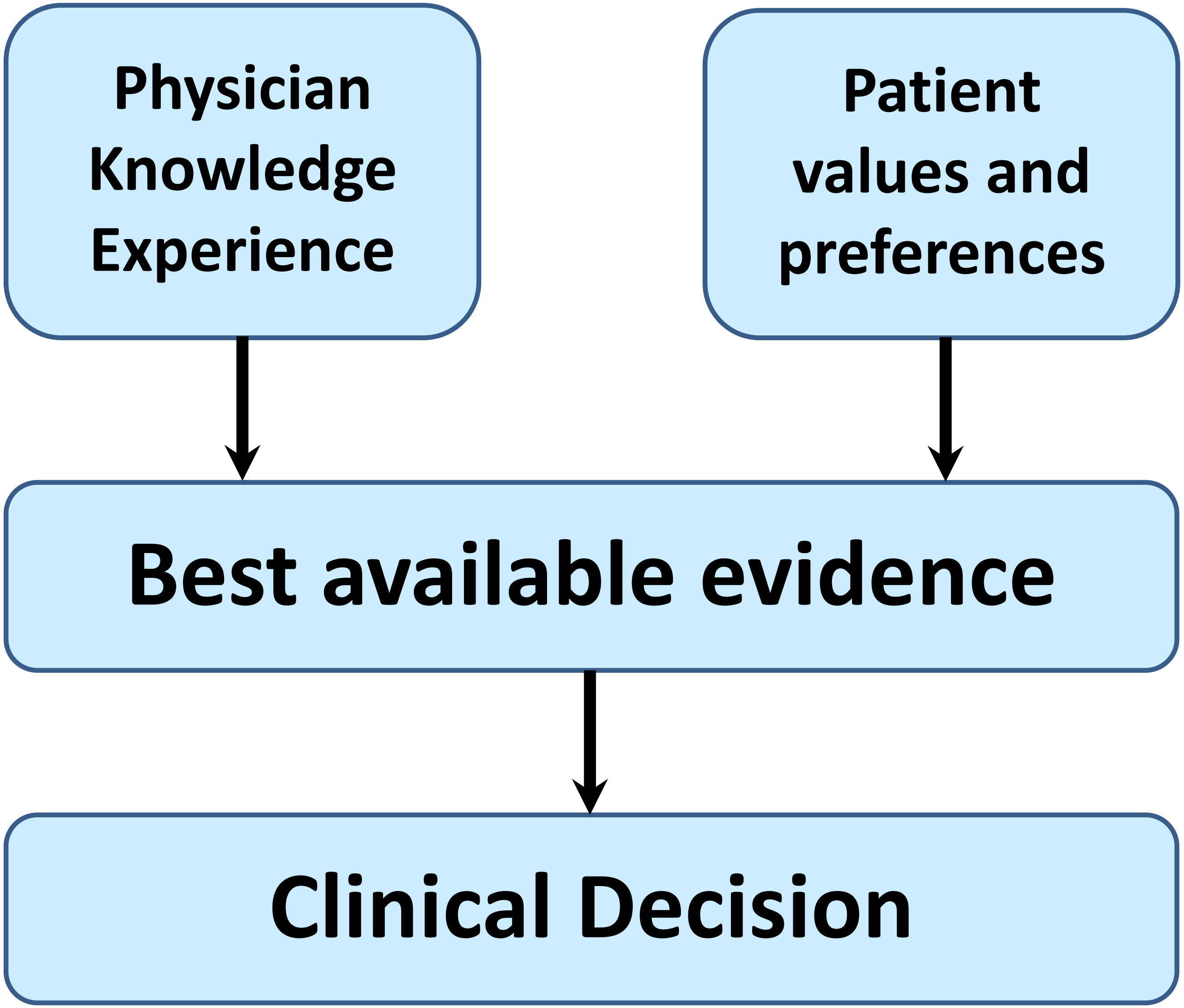
- There are perhaps 30,000 biomedical journals in the world, and they have grown steadily by 7% a year since the seventeenth century
- Yet only about 15% of medical interventions are supported by, solid scientific evidence
- This is partly because only 1% of the articles in medical journals are scientifically sound and partly because many treatments have never been assessed at all
- It is not enough to do the thing right; it is also necessary to do the right thing

BMJ. 1991 Oct 5;303(6806):798-9. http://www.ncbi.nlm.nih.gov/pubmed/1932964

Evidence Based Medicine

- Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients
- Evidence-based medicine is controversial not because people disagree about whether medical decisions ought to incorporate the best available evidence, but because they disagree about how narrowly evidence should be defined
- Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough

BMJ 1-13-96 http://www.bmj.com/content/312/7023/71.long
Acad Med. 2007 Mar;82(3):292-7 http://www.ncbi.nlm.nih.gov/pubmed/17327722



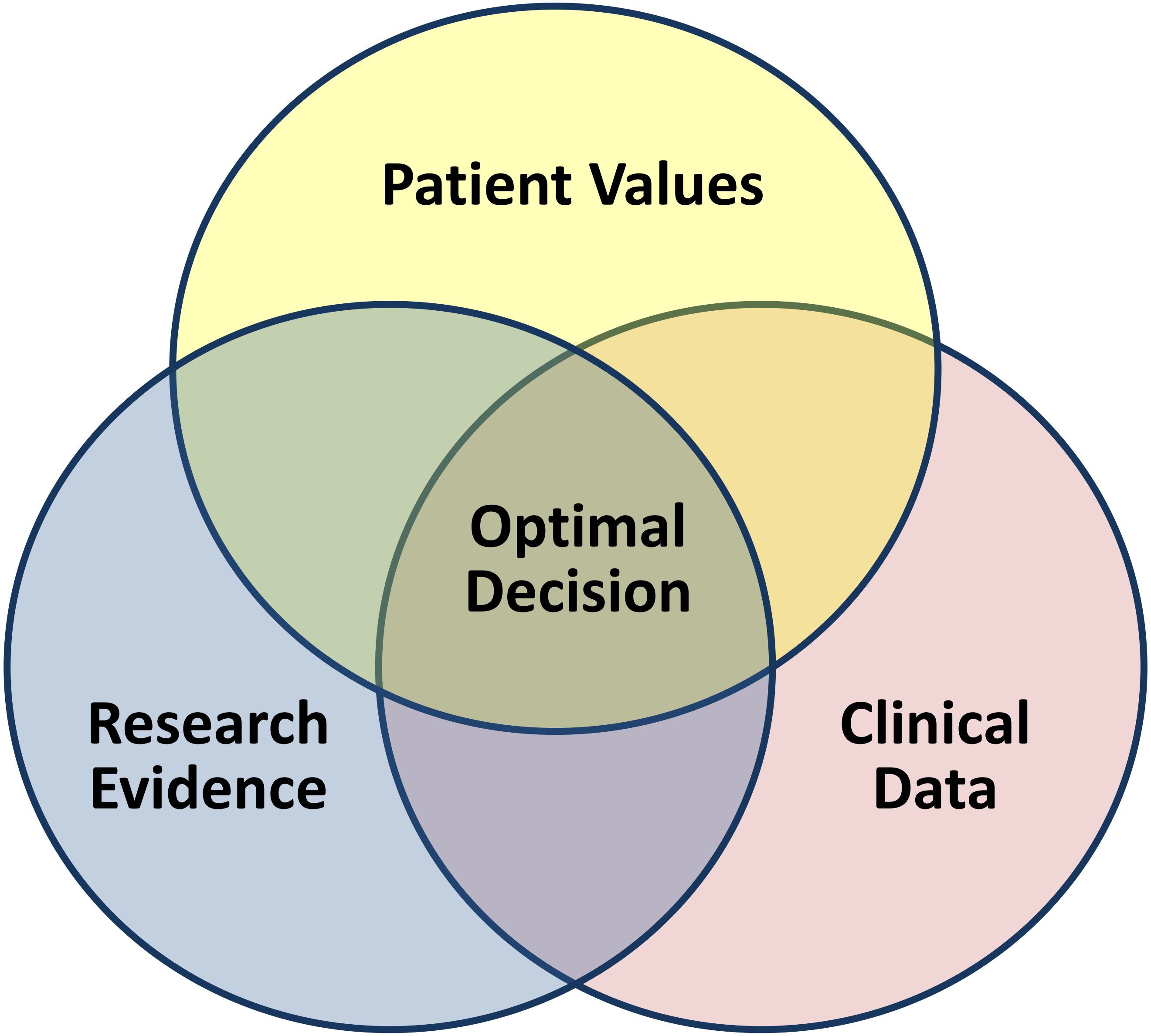
Physician
Knowledge
Experience

Patient
values and
preferences

Best available evidence

Clinical Decision

http://adc.bmj.com/content/90/8/837.long



Patient Values

Optimal Decision

Research Evidence

Clinical Data

https://med.mercer.edu/Library/mobile-ebm/glossary.htm

Evidence Based Medicine

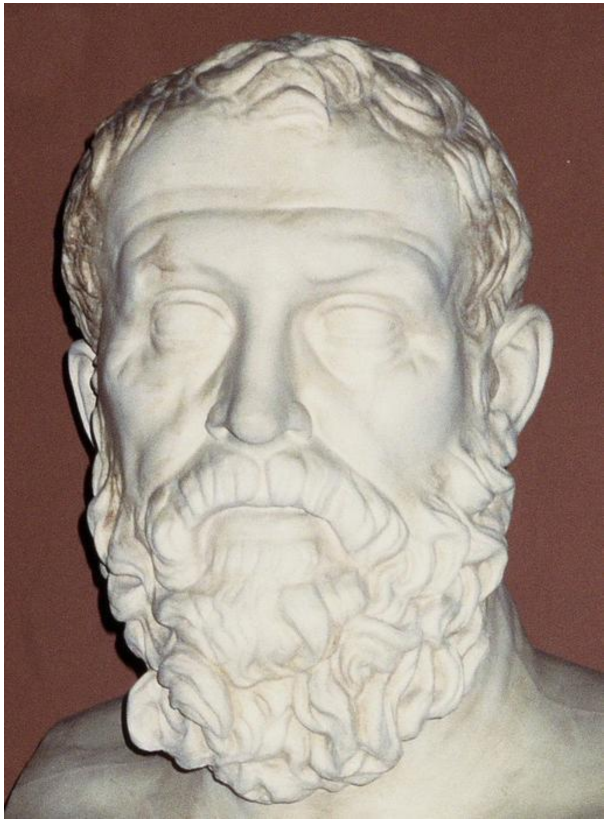
- Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient
- External clinical evidence can inform, but can never replace, individual clinical expertise, and it is this expertise that decides whether the external evidence applies to the individual patient at all and, if so, how it should be integrated into a clinical decision

BMJ 1-13-96 http://www.bmj.com/content/312/7023/71.long

The most basic of
Hippocratic principles:

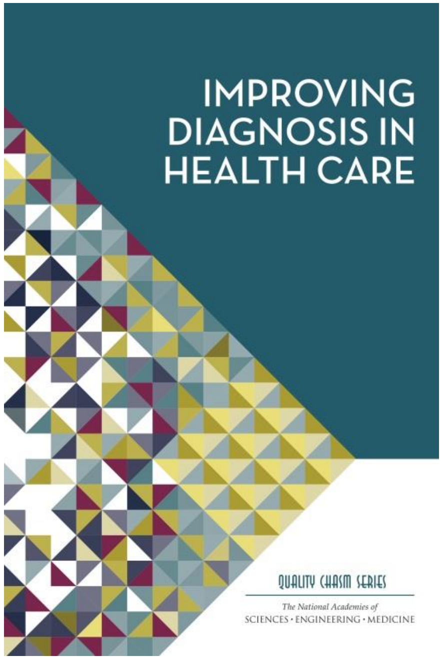
Primum non noncere

First, do no harm



Hippocrates
The Father of Medicine

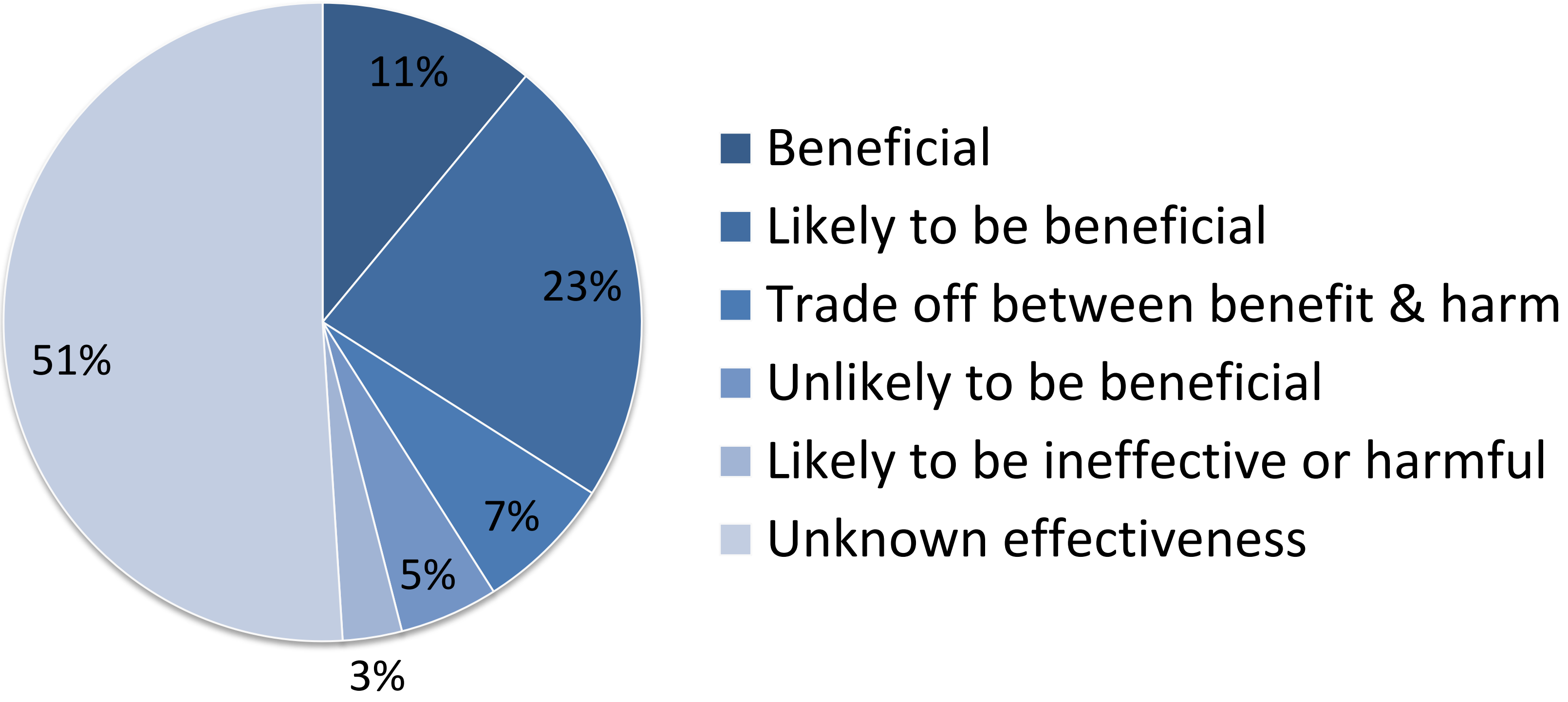
- Diagnostic Error: failure to
  - establish an accurate and timely explanation of the patient’s health problems
  - or communicate that explanation to the patient
- 5 percent of U.S. adults who seek outpatient care each year experience a diagnostic error


IMPROVING DIAGNOSIS IN HEALTH CARE
QUALITY CHASM SERIES
The National Academies of
SCIENCES · ENGINEERING · MEDICINE

- Most people will experience at least one diagnostic error in their lifetime, sometimes with devastating consequences
- Diagnostic errors contribute to approximately 10 percent of patient deaths

http://iom.nationalacademies.org/reports/2015/improving-diagnosis-in-healthcare

51% of Currently Used Treatments are of ‘Unknown Effectiveness’



http://clinicalevidence.bmj.com/ceweb/about/knowledge.jsp

Strength of Evidence

- Evidence is accumulating so rapidly that clinicians are often not able to assess and weigh the importance of the entire scope in detail
- It is, therefore, not surprising that several hierarchical systems for classifying the quality of evidence and generating recommendations have been created in order to guide the busy clinician for decision making and ultimately caring for patients

http://ccforum.com/content/10/5/232

EBM Hierarchy of Evidence

Strongest

Weakest

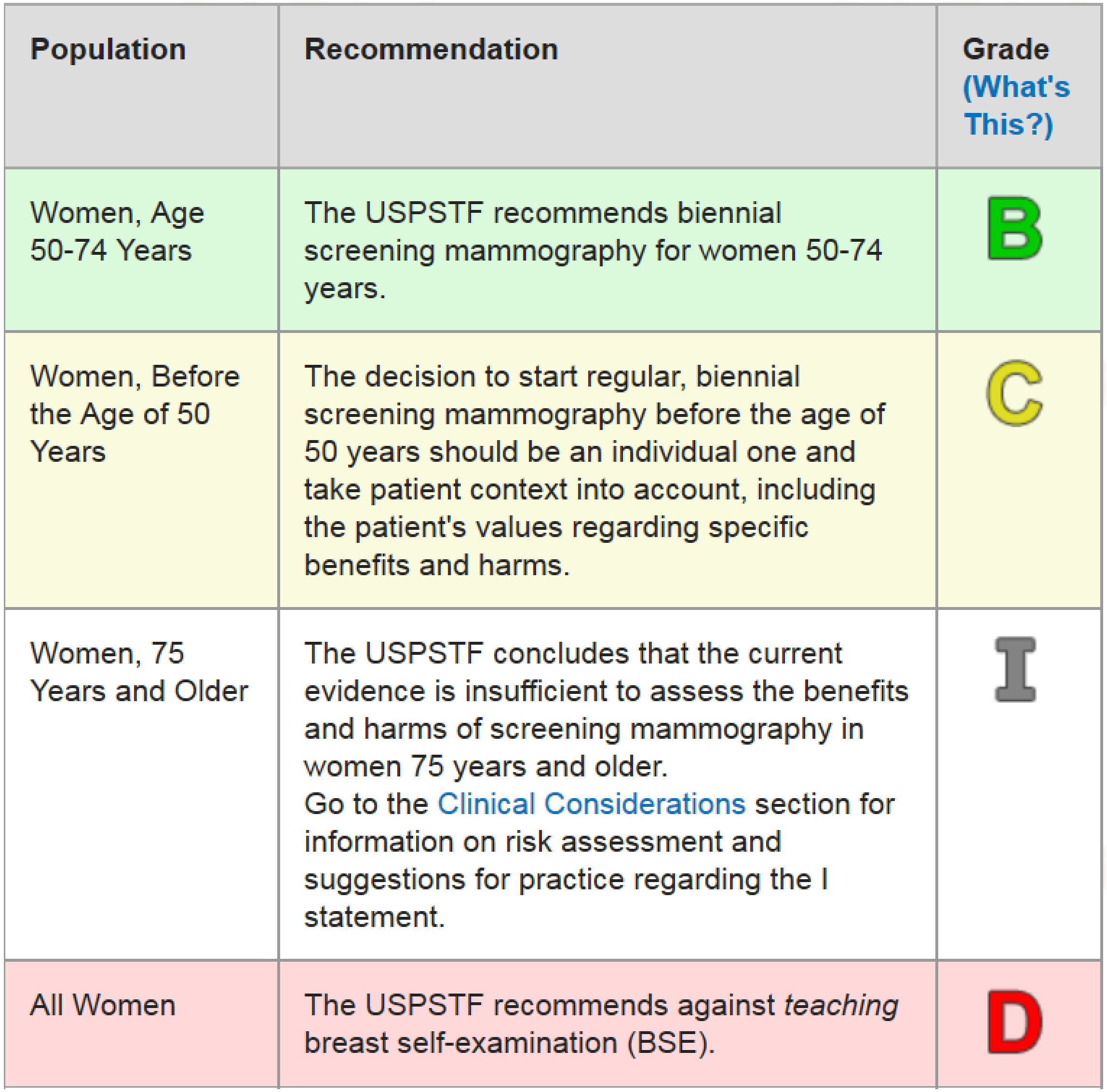
Level	Type of Evidence
1	Randomized clinical trial with an N of 1
2	Systematic review of randomized clinical trials
3	Randomized, placebo-controlled trial
4	Systematic review of observational outcomes research
5	Observational outcomes research
6	Physiologic studies
7	Unsystematic clinical observations

Acad Med. 2007 Mar;82(3):292-7 http://www.ncbi.nlm.nih.gov/pubmed/17327722

Grade	Strength of Recommendations
A	The USPSTF found good evidence that the service improves important health outcomes and concludes that benefits substantially outweigh harms
B	The USPSTF found at least fair evidence that the service improves important health outcomes and concludes that benefits outweigh harms
C	The USPSTF found at least fair evidence that the service can improve health outcomes but concludes that the balance of benefits and harms is too close to justify a general recommendation
D	The USPSTF found at least fair evidence that the service is ineffective or that harms outweigh benefits
I	Evidence that the service is effective is lacking, of poor quality, or conflicting and the balance of benefits and harms cannot be determined

http://www.uspreventiveservicestaskforce.org/3rduspstf/ratings.htm

2009 Breast Cancer Screening recommendations



Population	Recommendation	Grade (What's This?)
Women, Age 50-74 Years	The USPSTF recommends biennial screening mammography for women 50-74 years.	B
Women, Before the Age of 50 Years	The decision to start regular, biennial screening mammography before the age of 50 years should be an individual one and take patient context into account, including the patient's values regarding specific benefits and harms.	C
Women, 75 Years and Older	The USPSTF concludes that the current evidence is insufficient to assess the benefits and harms of screening mammography in women 75 years and older. Go to the Clinical Considerations section for information on risk assessment and suggestions for practice regarding the I statement.	I
All Women	The USPSTF recommends against *teaching* breast self-examination (BSE).	D

http://www.uspreventiveservicestaskforce.org/uspstf/uspsbrca.htm

Criticisms of Evidence Based Medicine

- Evidence-based medicine provides no theoretical or logical justification for the premise that epidemiologic and statistical information is inherently superior to other types of information
- Conclusions drawn from evidence-based medicine's population-level data have limited relevance to treatment decisions for individual patients

Acad Med. 2007 Mar;82(3):292-7 http://www.ncbi.nlm.nih.gov/pubmed/17327722

Criticisms of Evidence Based Medicine

- Health care managers and administrators often co-opt evidence-based medicine's methods and rhetoric to decrease costs or inappropriately restrict clinical practice
- No logical or empirical proof has demonstrated that following the tenets of evidence-based medicine leads to better patient outcomes

Acad Med. 2007 Mar;82(3):292-7 http://www.ncbi.nlm.nih.gov/pubmed/17327722

Who Pays for Research?

- Pharmaceutical companies, who are major funders of research worldwide, are interested in products that will find a market, irrespective of the causes and distribution of ill health within populations
- There is no incentive for these companies to fund small studies with disadvantaged groups because, even if the research identifies effective drug treatments, the market will be small and therefore unlikely to be profitable in the long run

J Med Ethics. 2004 Apr;30(2):141-5. http://www.ncbi.nlm.nih.gov/pubmed/15082806

Who Pays for Research?

- One third of the trials published in the BMJ and between two thirds and three quarters published in the major North American journals were funded by the pharmaceutical industry
- *“Journals have devolved into information laundering operations for the pharmaceutical industry”*

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Evaluation of Conflicts of Interest in Economic Analyses of New Drugs Used in Oncology

	Industry Sponsored Research	
	Yes	No
Favorable Qualitative Conclusion	19 a	b 15
Unfavorable Qualitative Conclusion	c 1	d 9

Odds Ratio = **a∗d** / **b∗c** = 11.4

JAMA. 1999 Oct 20;282(15):1453-7. http://www.ncbi.nlm.nih.gov/pubmed/10535436

The Price Tag on Progress

- A NEJM Perspective editorial looked at the chemotherapy costs for colorectal cancer
- The Mayo Clinic treatment regimen gives a median survival time of 21 months at a cost of $63 using generic drugs
- The new ‘improved’ drug regimen increases the median survival time by 1.7 months, at a cost of $30,790
- If all of the individuals with colorectal cancer were switched to this new drug regimen, it would result in a societal cost of $666,000,000

N Engl J Med. 2004 Jul 22;351(4):317-9. http://www.ncbi.nlm.nih.gov/pubmed/15269308

Epidemiology, Disease Progression, and Economic Burden of Colorectal Cancer

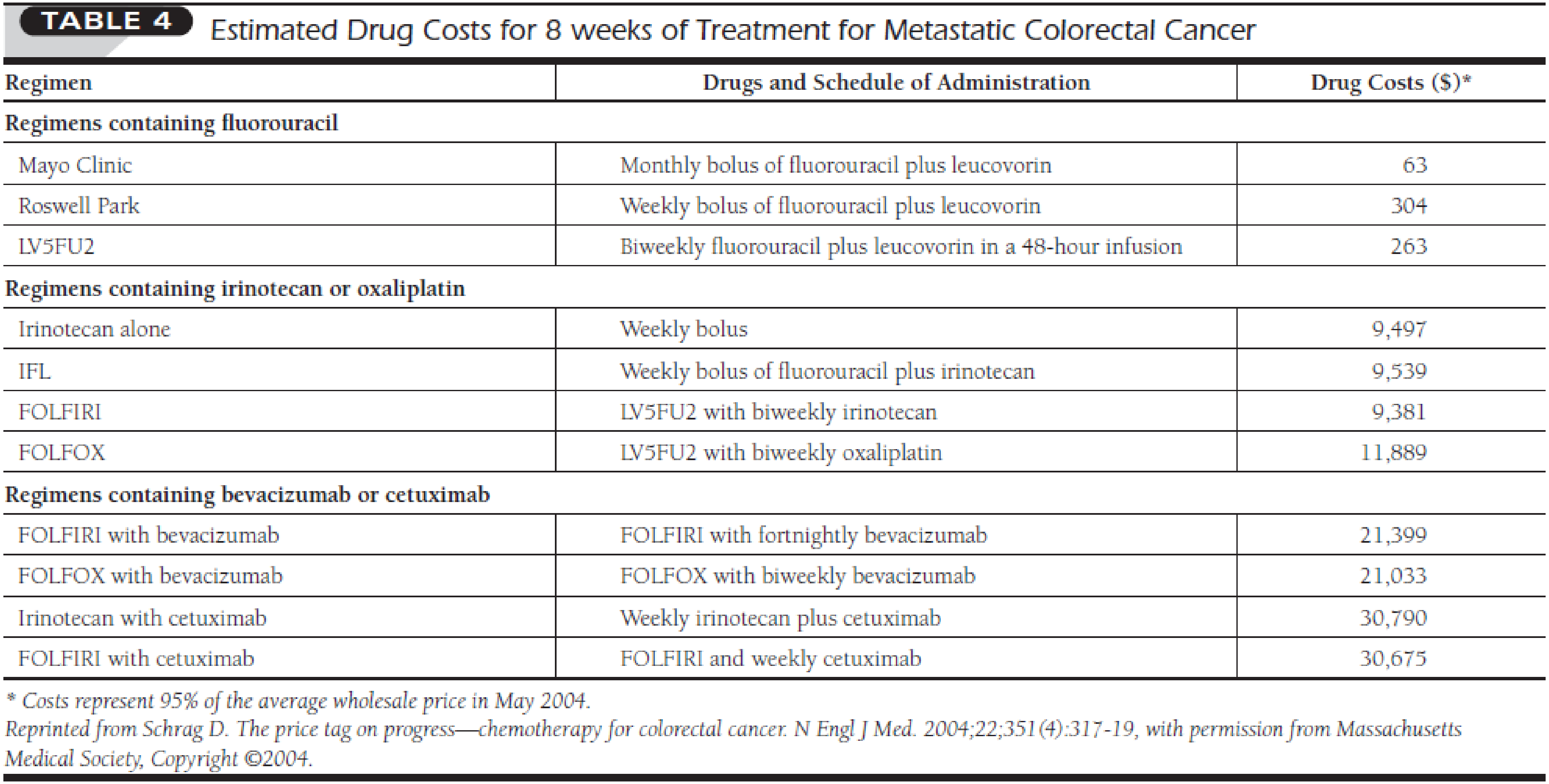


TABLE 4 Estimated Drug Costs for 8 weeks of Treatment for Metastatic Colorectal Cancer

Regimen	Drugs and Schedule of Administration	Drug Costs ($)*
Regimens containing fluorouracil		
Mayo Clinic	Monthly bolus of fluorouracil plus leucovorin	63
Roswell Park	Weekly bolus of fluorouracil plus leucovorin	304
LV5FU2	Biweekly fluorouracil plus leucovorin in a 48-hour infusion	263
Regimens containing irinotecan or oxaliplatin		
Irinotecan alone	Weekly bolus	9,497
IFL	Weekly bolus of fluorouracil plus irinotecan	9,539
FOLFIRI	LV5FU2 with biweekly irinotecan	9,381
FOLFOX	LV5FU2 with biweekly oxaliplatin	11,889
Regimens containing bevacizumab or cetuximab		
FOLFIRI with bevacizumab	FOLFIRI with fortnightly bevacizumab	21,399
FOLFOX with bevacizumab	FOLFOX with biweekly bevacizumab	21,033
Irinotecan with cetuximab	Weekly irinotecan plus cetuximab	30,790
FOLFIRI with cetuximab	FOLFIRI and weekly cetuximab	30,675

* *Costs represent 95% of the average wholesale price in May 2004.*

Reprinted from Schrag D. The price tag on progress—chemotherapy for colorectal cancer. N Engl J Med. 2004;22;351(4):317-19, with permission from Massachusetts Medical Society, Copyright ©2004.

J Manag Care Pharm. 2007 Aug http://www.ncbi.nlm.nih.gov/pubmed/17713990
www.amcp.org/data/jmcp/Pages%205-18.pdf

Cost-Effectiveness Ratio

A CE ratio can be interpreted as the "price" of accepting a new intervention

$$\text{CE ratio} = \frac{\text{Cost with intervention} - \text{Cost without intervention}}{\text{Outcome with intervention} - \text{Outcome without intervention}}$$

$$\frac{\$30{,}790 - \$63}{22.7 \text{ months} - 21 \text{ months}} = \$18{,}075 \text{ / month}$$

A medical intervention is considered cost-effective when the intervention provides a health benefit at an "acceptable cost"

www.businessgrouphealth.org/preventive/background/part1.pdf

Drug Company Tactics

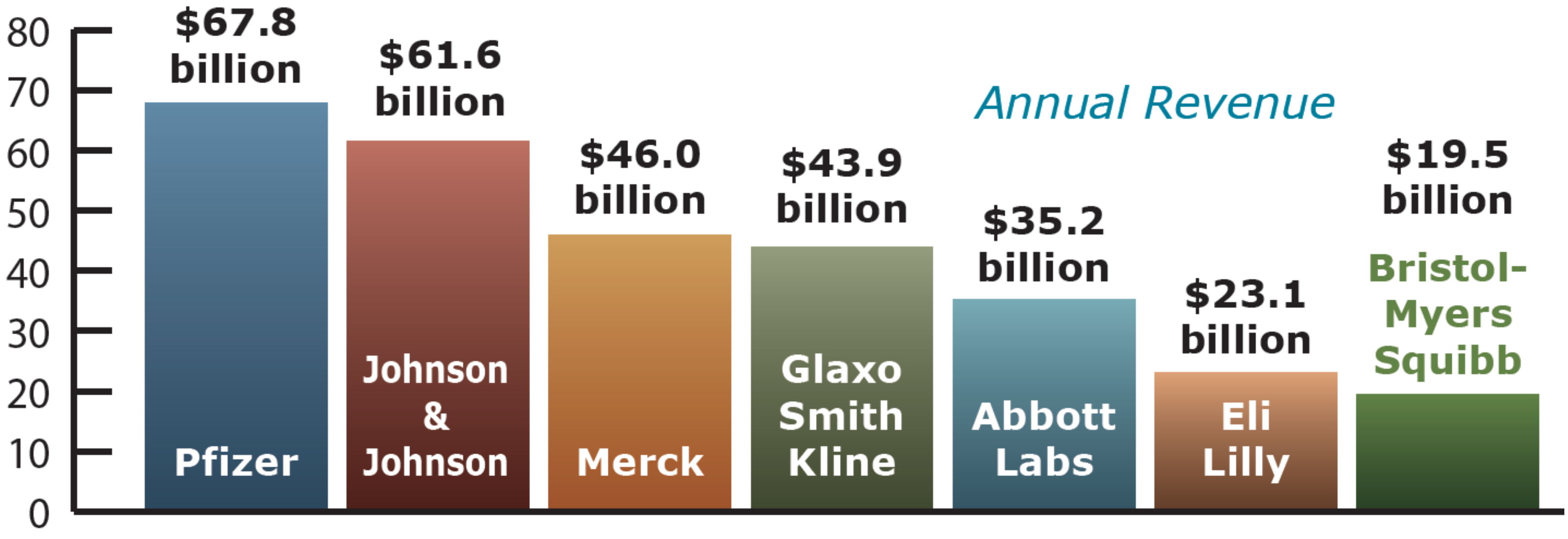
- Conduct a trial of your drug against a treatment known to be inferior
- Trial your drugs against too low a dose of a competitor drug
- Do multicentre trials and select for publication results from centres that are favourable
- Conduct subgroup analyses and select for publication those that are favourable
- Present results that are most likely to impress – for example, reduction in relative rather than absolute risk

PLoS Med. 2005 May;2(5):e138. 2005 May 17. www.ncbi.nlm.nih.gov/pubmed/15916457

Steven Nissen – Escape Fire

https://www.youtube.com/watch?v=JYL--jULtdM&hd=1

The Cost of Doing Business



SOURCE: CNN Money (http://money.cnn.com). Global 500 – annual ranking of the world's largest corporations, 2011.

How painful was the $3 Billion fine that Glaxo Smith Kline had to pay?

Disease Mongering

- Take a common symptom that could mean anything and make it sound as if it is a sign of a serious disease
- Define as large a proportion of the population as possible as suffering from the ‘disease’
- Selective use of statistics to exaggerate the benefits of treatment
- Get the right spin doctors

PLoS Med. 2006 Apr;3(4):e178 http://www.ncbi.nlm.nih.gov/pubmed/16597176

Randomized Clinical Trials

- Randomized Clinical trials work best on simple interventions and homogenous populations
- Randomised controlled trials carried out in specialised units by expert care givers, designed to determine whether an intervention does more good than harm under ideal conditions, cannot tell us how experimental treatments will fare in general use, nor can they identify rare side effects

J Med Ethics. 2004 Apr;30(2):151-5. http://www.ncbi.nlm.nih.gov/pubmed/15082808
BMJ. 1997 December 20 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2128012

Randomized Clinical Trials

- In other words, although an RCT may be *internally* valid, this does not necessarily mean that it is also *externally* valid
- Where the RCT has been most successful is in those cases in which both internal and external validity are high – eg, in situations dealing with homogeneous systems such as the evaluation of pharmaceutical treatment interventions
- In the case of pharmaceutical interventions, there is little argument that what counts in EBP is evidence from well conducted RCTs

J Manipulative Physiol Ther. June 2001 http://www.ncbi.nlm.nih.gov/pubmed/11416828

Ethics of Evidence Based Medicine

- Medical research, particularly the “gold standard” RCT, is much more common in the area of drug interventions
- This is partly because the pharmaceutical industry is a major source of research funding
- Other interventions, including the more complex interventions often required in primary care, are less likely to have received the attention of well funded research, and thus less likely to have robust evidence of clinical and cost effectiveness

J Med Ethics. 2004 Apr;30(2):151-5. http://www.ncbi.nlm.nih.gov/pubmed/15082808

Ethics of Evidence Based Medicine

- Most RCTs use outcome measures that have been designed to be easily quantifiable and amenable to statistical analysis
- **Research evidence is evidence about populations of patients rather than individuals**
- Significant risk reduction at a population level may not be so important for an individual when weighed against other considerations in that individual’s life

J Med Ethics. 2004 Apr;30(2):151-5. http://www.ncbi.nlm.nih.gov/pubmed/15082808

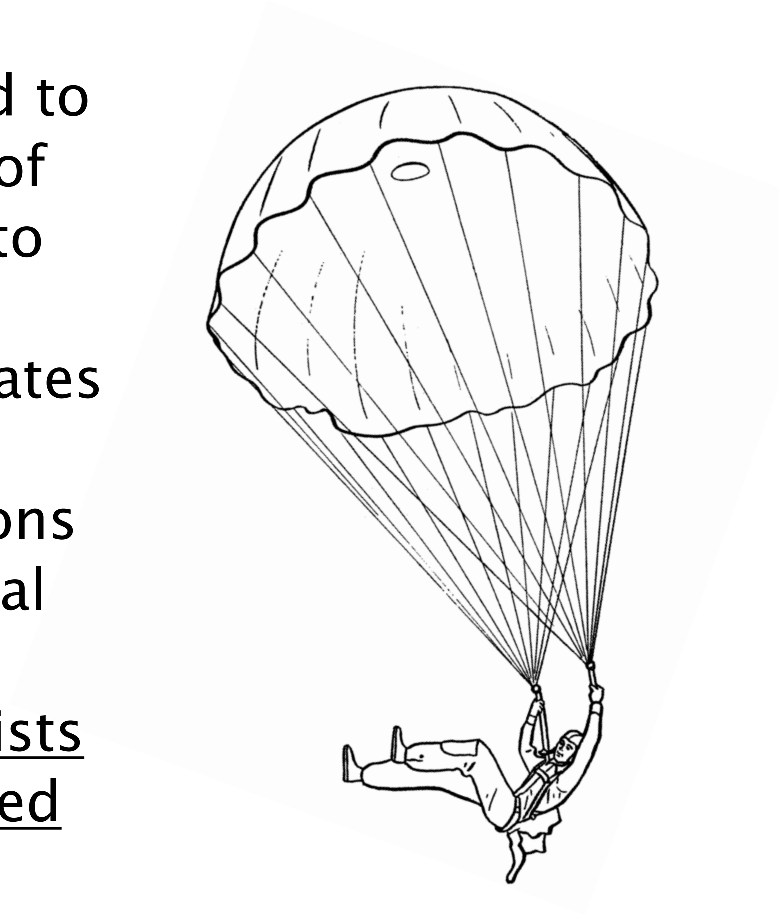
Health Promotion Research

- We argue that the ‘gold standard’ randomized controlled trial (RCT) should rarely, if ever, be used to evaluate health promotion
- Health promotion should be concerned with empowerment and must involve the participation of the public, clients and patients

http://www.amazon.com/Health-Promotion-Professor-Keith-Tones/dp/0761974490

Common Sense vs. the RCT

As with many interventions intended to prevent ill health, the effectiveness of parachutes has not been subjected to rigorous evaluation by using randomised controlled trials. Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute.



Parachute use to prevent death and major trauma related to gravitational challenge
BMJ 12-20-03 http://www.ncbi.nlm.nih.gov/pubmed/14684649

Epidemiology

- The word epidemiology has its roots in the study of what befalls a population:
  - *Epi* – on or upon
  - *demos* – people
  - *logos* – the study of
- The basic premise of epidemiology is that disease does not occur randomly but in patterns that reflect underlying factors that can be determined

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Definitions

- Endemic refers to the constant presence of a disease or infectious agent in a population within a geographic area
- Epidemic refers to a sudden increase in the number of cases of a disease above what is normally expected in that population in that area
- Outbreak carries the same definition of epidemic, but is often used for a more limited geographic area
- Pandemic refers to an epidemic that has spread over several countries or continents, usually affecting a large number of people

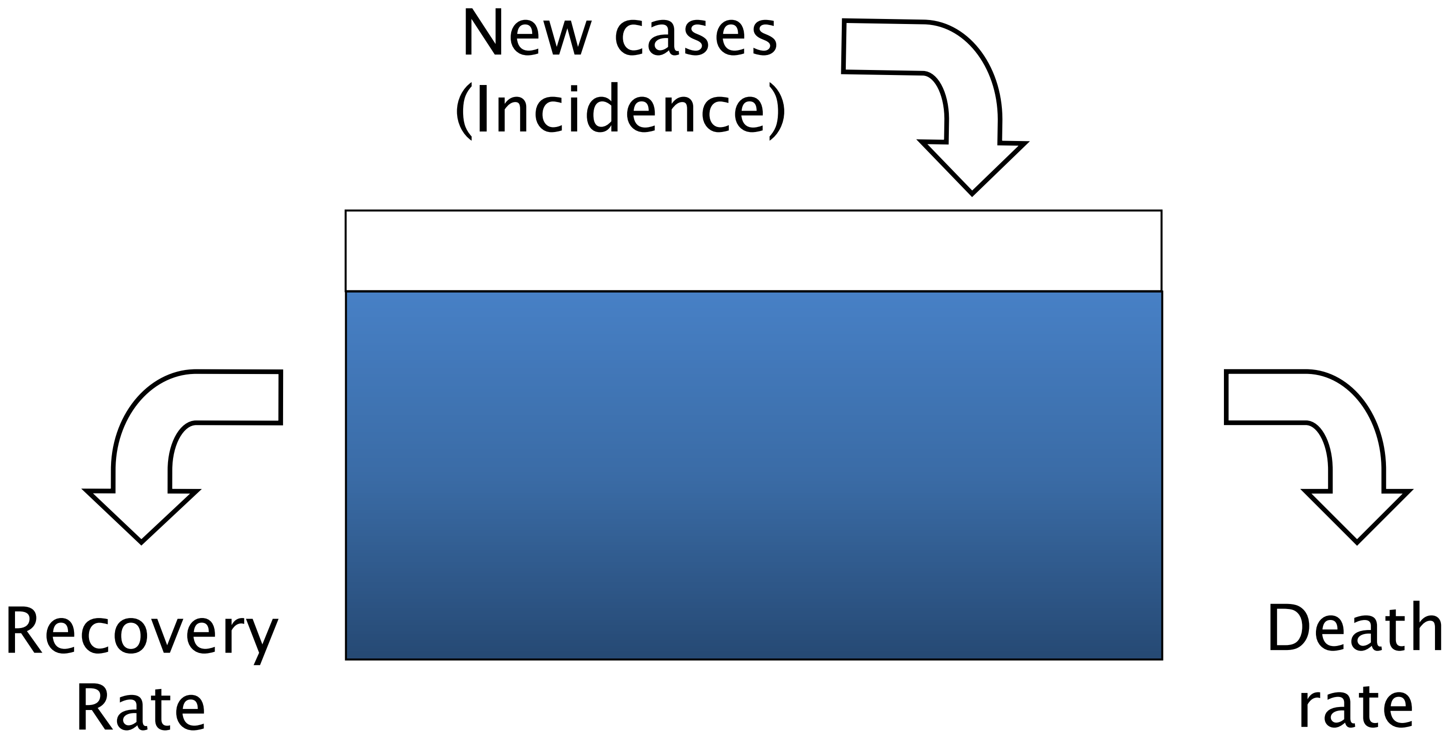
Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Definitions

- Morbidity is the presence of disease, injury, or disability in the population
- Morbidity frequency is measured by:
    - Incidence – the number of persons in a population who become ill
    - Prevalence – the number of persons who are ill at a given time

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

The Prevalence “Pool”



New cases
(Incidence)

Recovery
Rate

Death
rate

Definitions

- Mortality measures the number of deaths in a defined population during a specified interval
- Morbidity and mortality measures are often the same mathematically; it's just a matter of what you choose to measure, illness or death

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

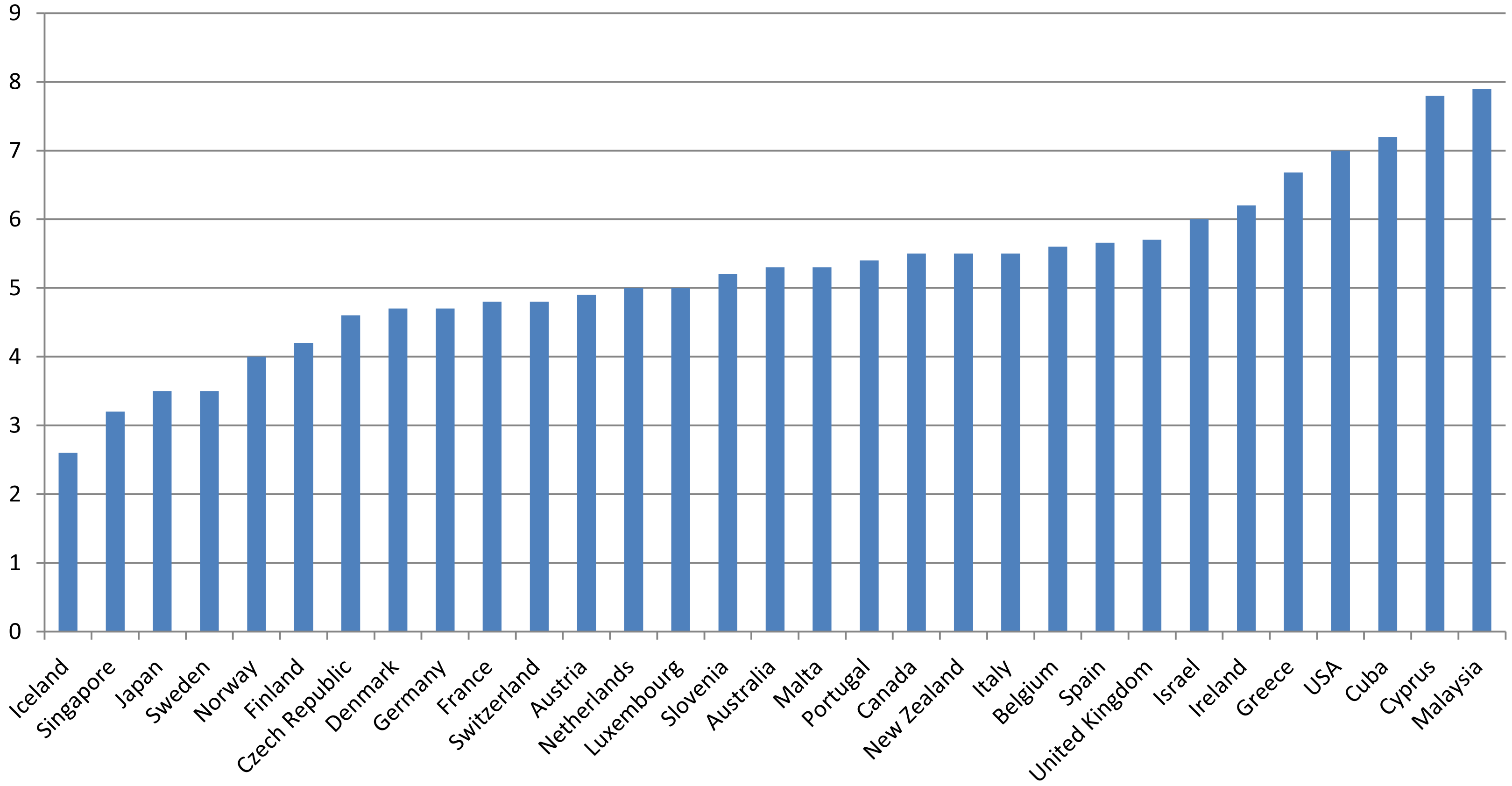
Important Mortality Rates

Rate	Definition	Example
Crude death rate	$\frac{\text{Total number of deaths}}{\text{Midyear population}}$	8.0 / 1,000
Cause-specific death rate	$\frac{\text{Number of deaths due to a specific cause}}{\text{Midyear population}}$	26 / 100,000
Infant mortality rate	$\frac{\text{Number of deaths < 1 year of age}}{\text{Midyear population}}$	7.0 / 1,000

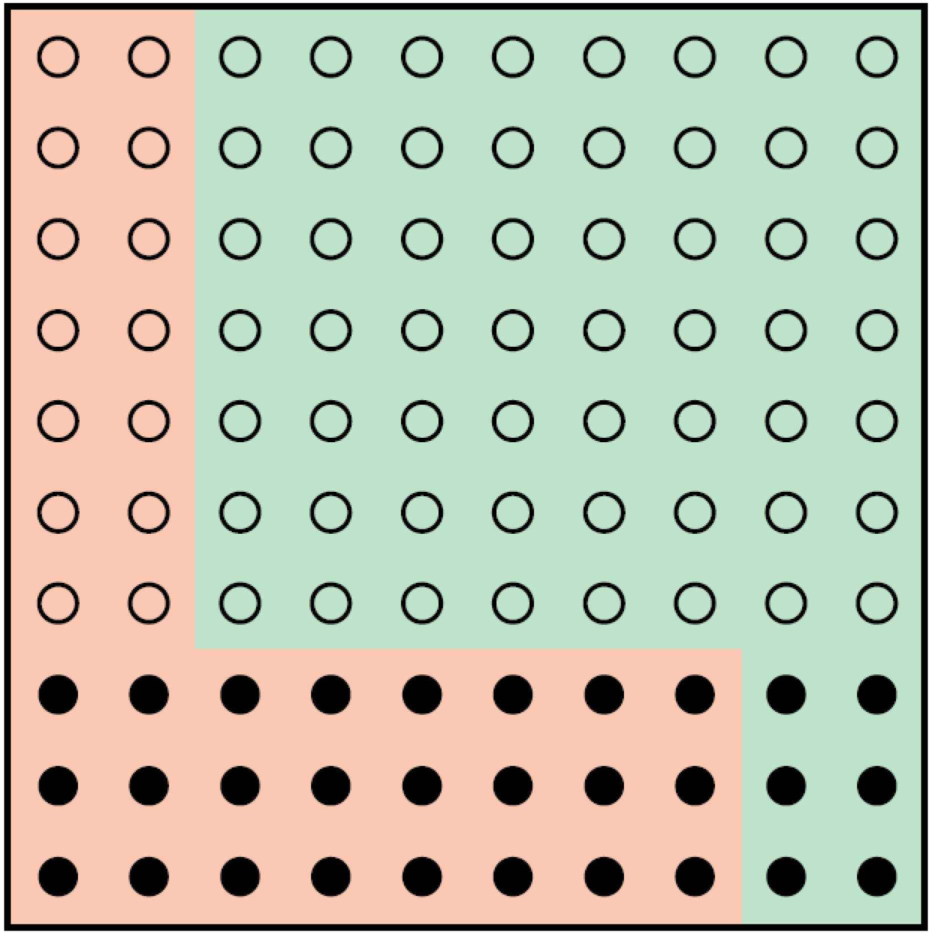
Because rates put disease frequency in the perspective of the size of the population, rates are particularly useful for comparing disease frequency in different locations, at different times, or among groups with different sized populations

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Infant Mortality Rates (per 1,000 live births)



Understanding Sensitivity and Specificity with the Right Side of the Brain



○ = Disease absent

● = Disease present

🟩 = Test negative

🟥 = Test positive

🟥● = True positive

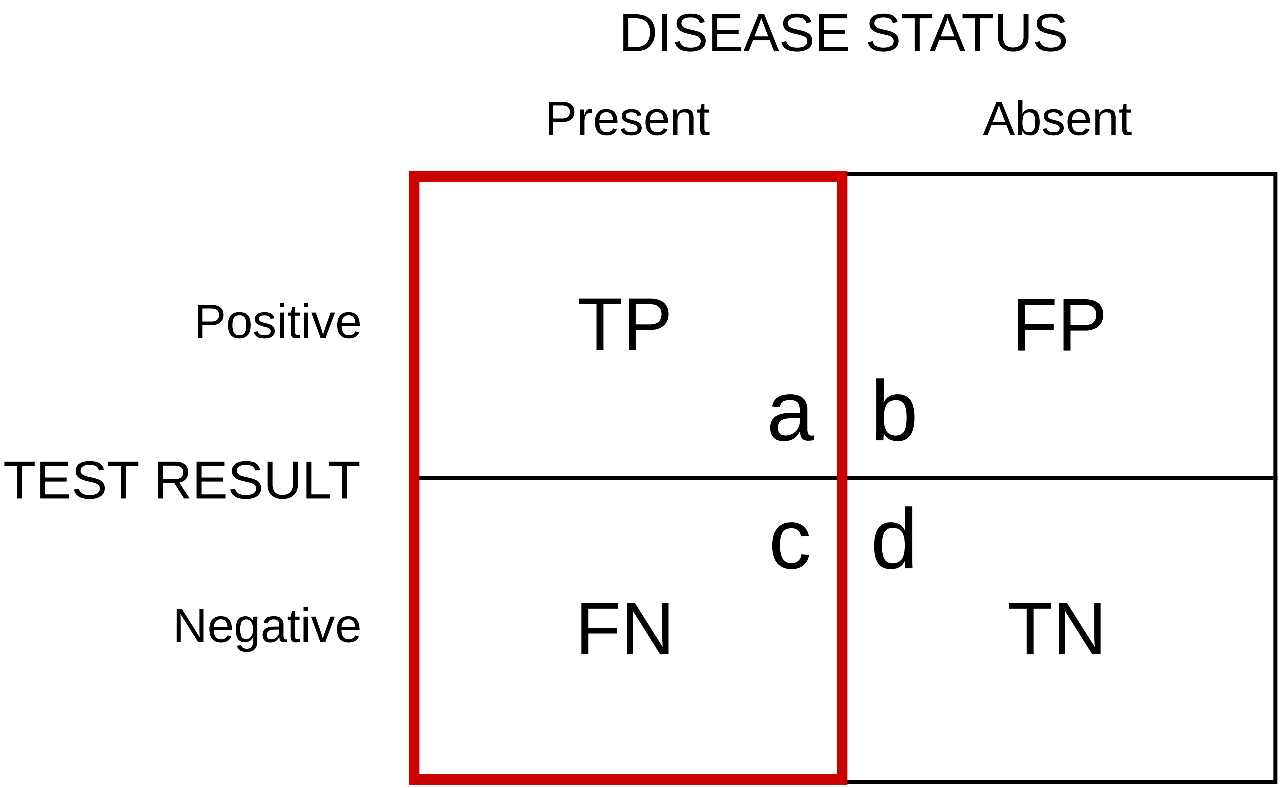
🟩○ = True negative

🟥○ = False positive

🟩● = False negative

BMJ 9-27-03 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/

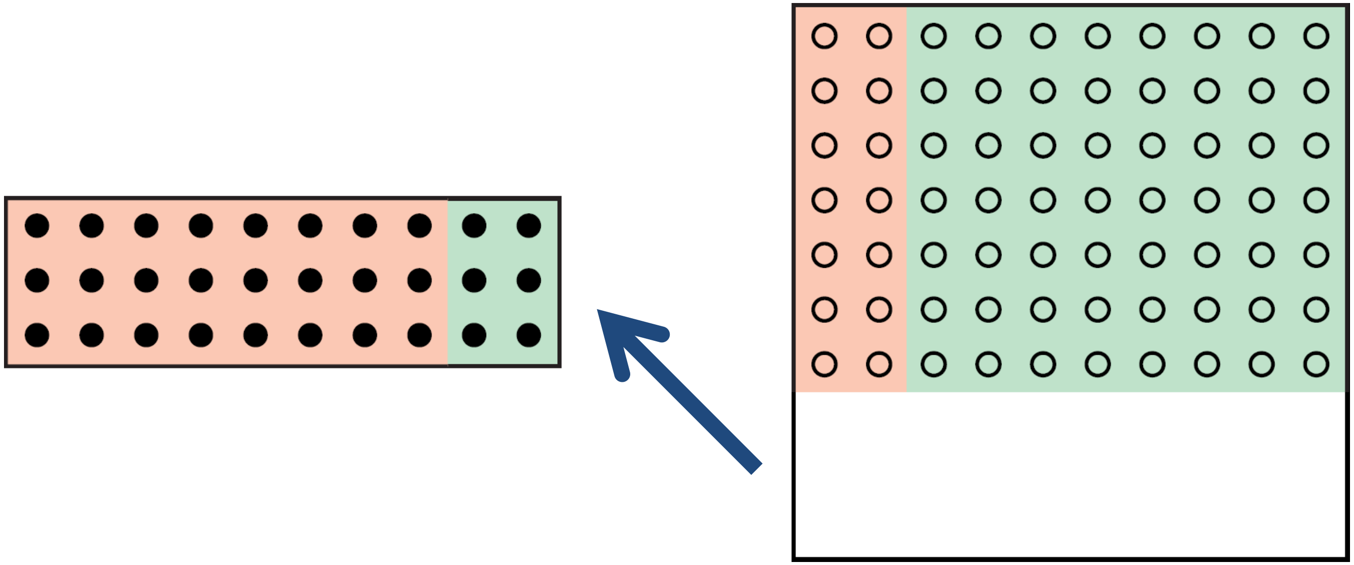
Sensitivity = TP / (TP+FN)



		DISEASE STATUS	
		Present	Absent
TEST RESULT	Positive	TP a	b FP
	Negative	c FN	d TN

Sensitivity is the proportion of people **with** the disease who will have a **positive** test result

Sensitivity = TP / (TP+FN)



24 / (24+6) = 80% sensitivity

● / ● + ●

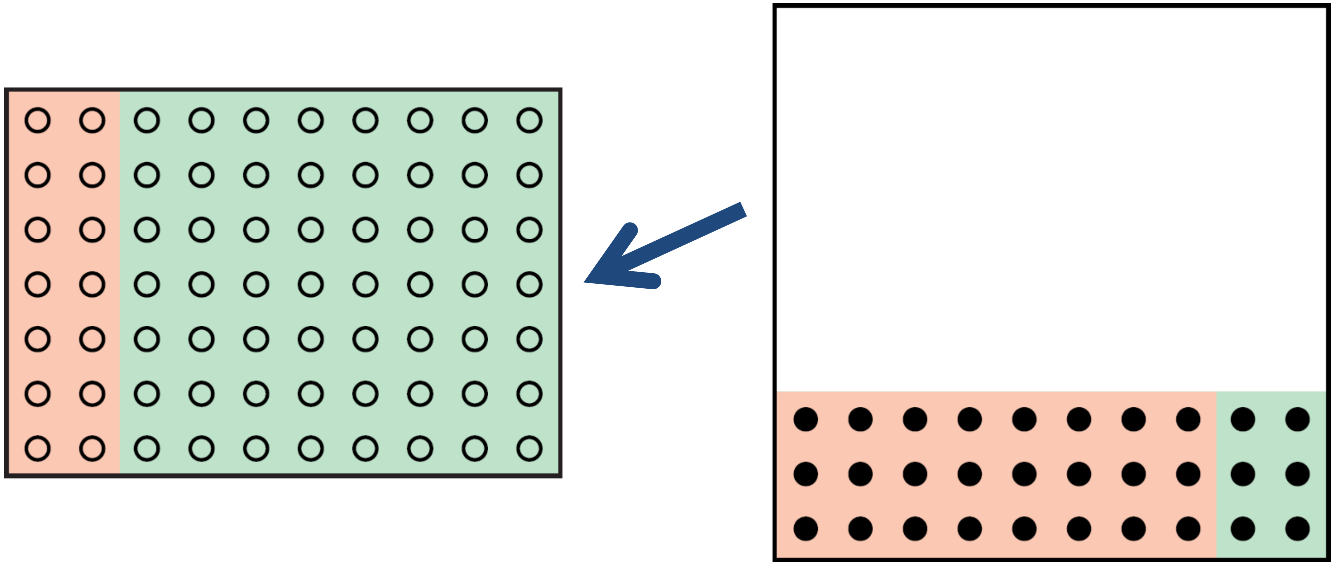
BMJ 9-27-03 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/

Specificity = TN / (TN+FP)

		DISEASE STATUS	
		Present	Absent
TEST RESULT	Positive	TP a	b FP
	Negative	c FN	d TN

Specificity is the proportion of people **without** the disease who will have a **negative** test result

Specificity = TN / (TN+FP)



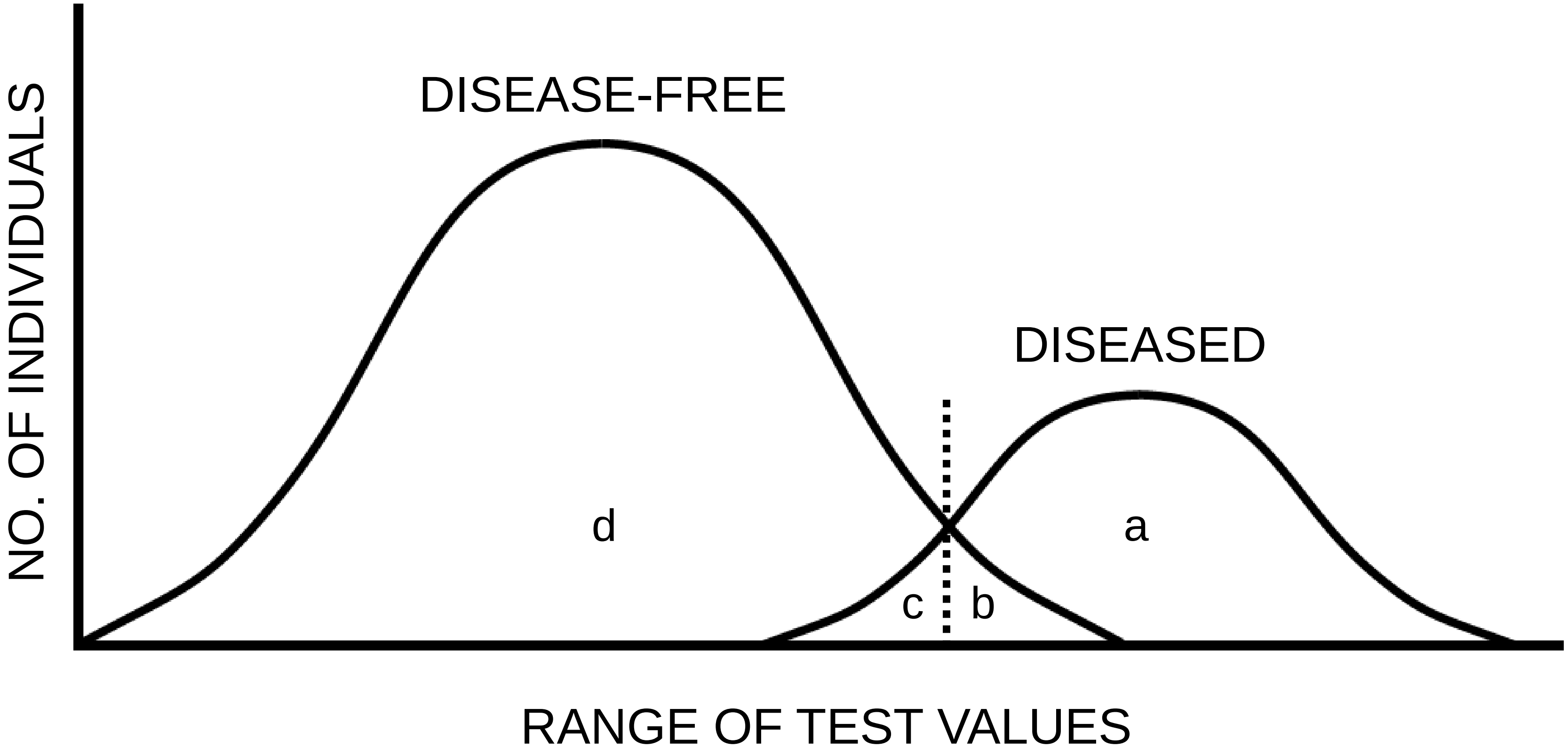
56 / (56+14) = 80% specificity

🟩 / 🟩 + 🟧

BMJ 9-27-03 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/

Sensitivity vs. Specificity

Sensitivity	Specificity
Calculated only from individuals who have the disease	Calculated only from individuals who do not have the disease
'True Positive' rate	'True Negative' rate
The greater the test's sensitivity, the lower the false negative rate (FN = Type II error)	The greater the test's specificity, the lower the false positive rate (FP = Type I error)
Sn-N-Out If a test has high **sensitivity**, a **negative** result rules **out** the disease	Sp-P-In If a test has high **specificity**, a **positive** result rules **in** the disease

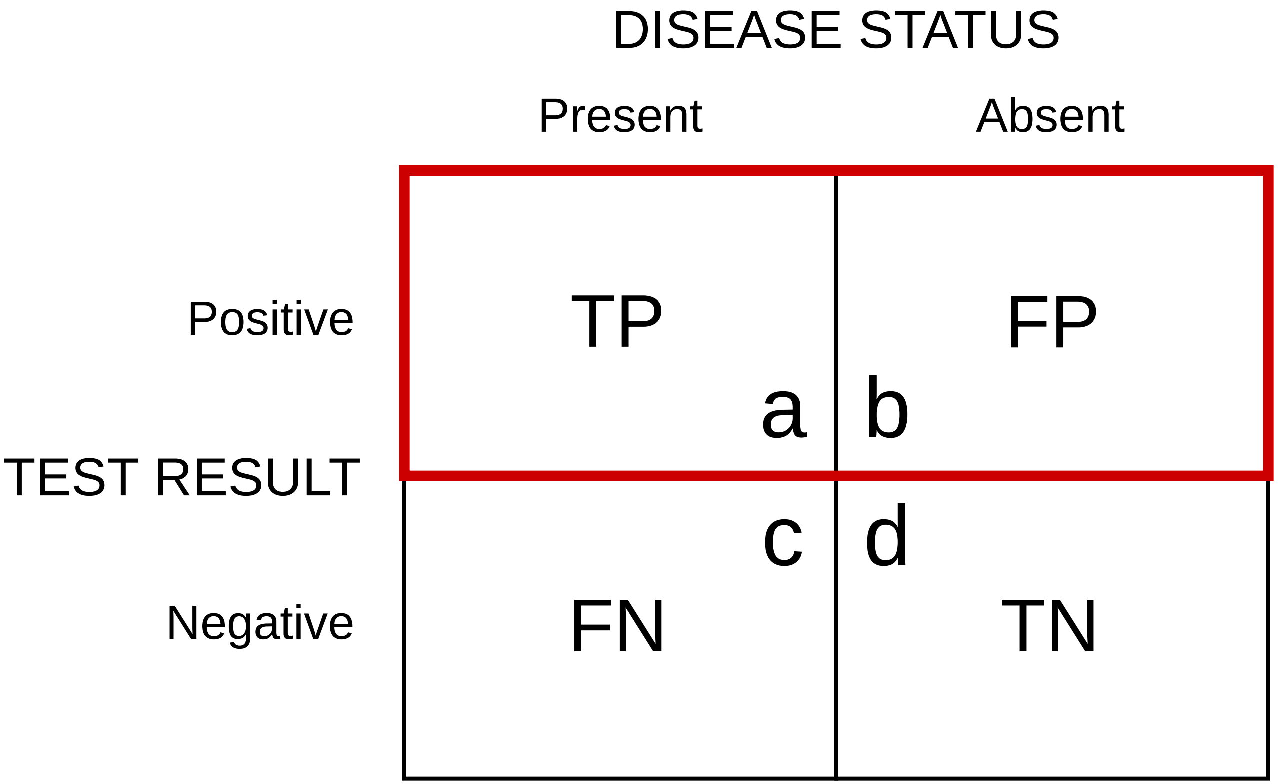


a = True Positive
b = False Positive

c = False Negative
d = True Negative

Can Med Assoc J. 1981 Mar 15;124(6):703–10. www.ncbi.nlm.nih.gov/pubmed/7471014

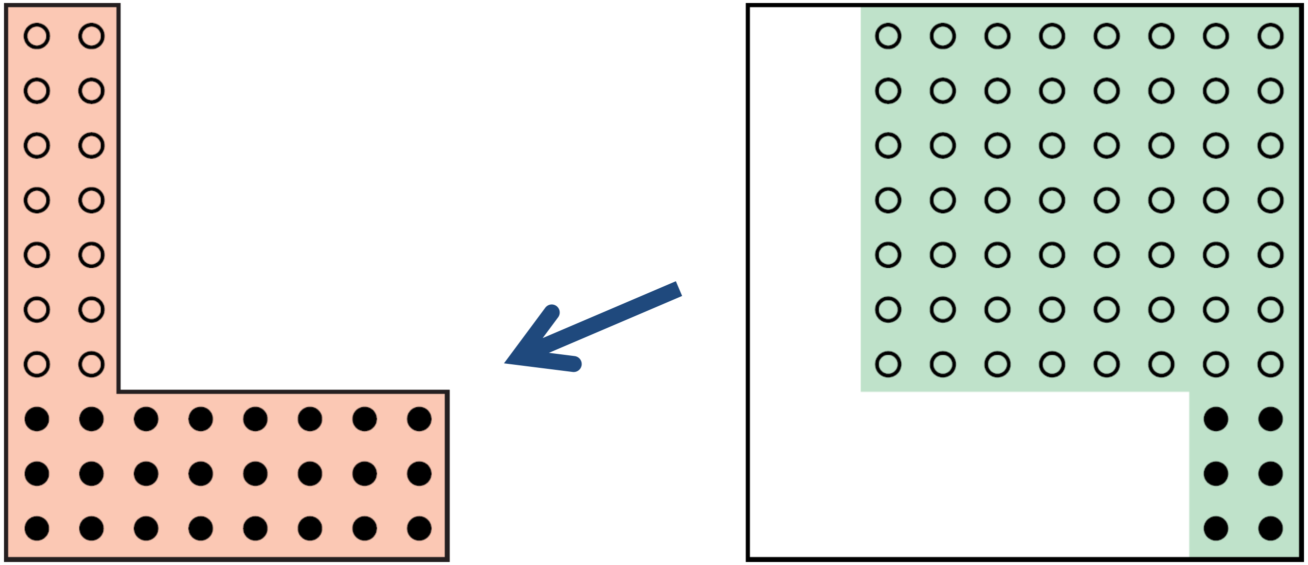
PPV = TP / (TP+FP)



		DISEASE STATUS	
		Present	Absent
TEST RESULT	Positive	TP a	FP b
	Negative	c FN	d TN

Positive Predictive Value (PPV) is the proportion of people with a **positive** test result who actually **have** the disease

PPV = TP / (TP+FP)



24 / (24+14) = 63% PPV

● / ● + ○

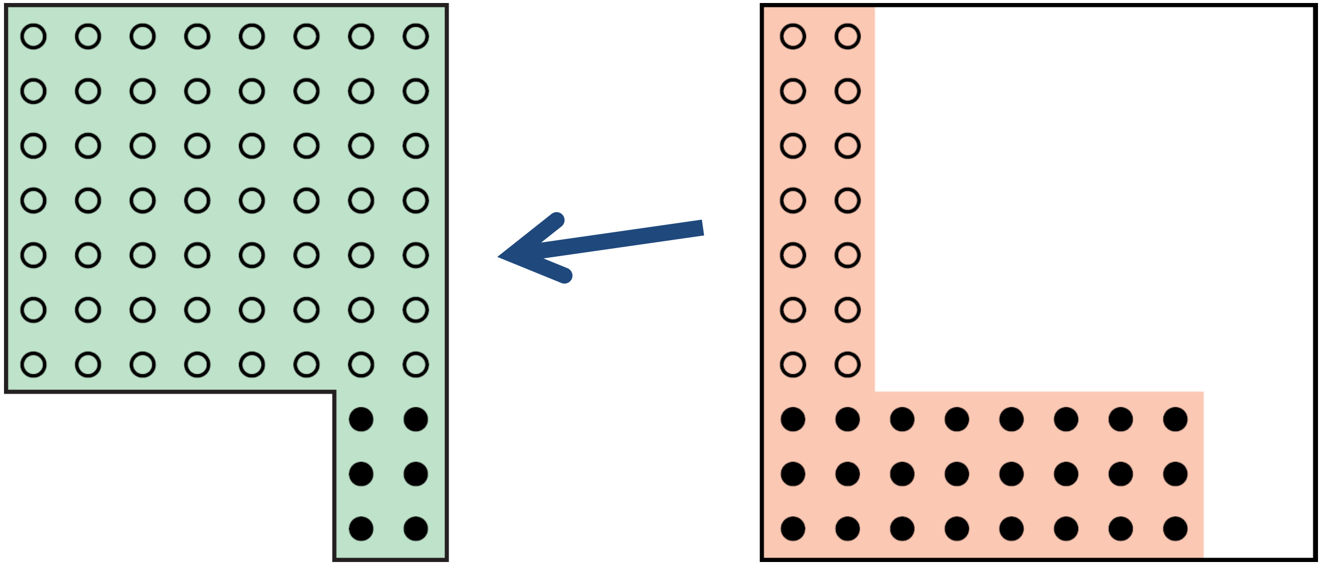
BMJ 9-27-03 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/

NPV = TN / (TN+FN)

		DISEASE STATUS	
		Present	Absent
TEST RESULT	Positive	TP a	b FP
	Negative	c FN	d TN

Negative Predictive Value (NPV) is the proportion of people with a **negative** test result who do **not** have the disease

NPV = TN / (TN+FN)

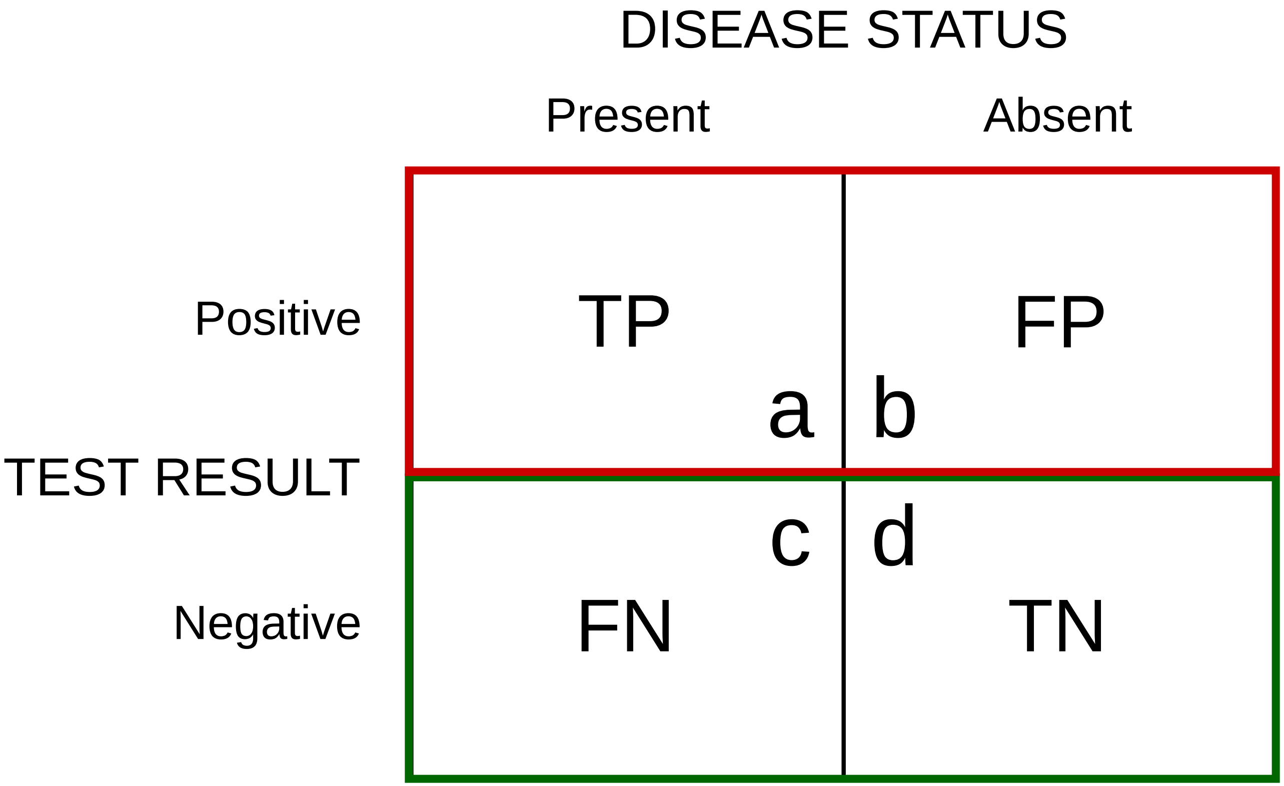


56 / (56+6) = 90% NPV

○ / ○ + ●

BMJ 9-27-03 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC200804/

Risk Difference = PPV – NPV



DISEASE STATUS

TEST RESULT

	Present	Absent
Positive	TP a	b FP
Negative	c FN	d TN

Risk Difference, also known as Absolute Risk Reduction (ARR), is the difference between two groups, typically between the exposed experimental group vs. the unexposed control group

Number Needed to Treat

- Number Needed to treat is the inverse of absolute risk reduction (ARR):

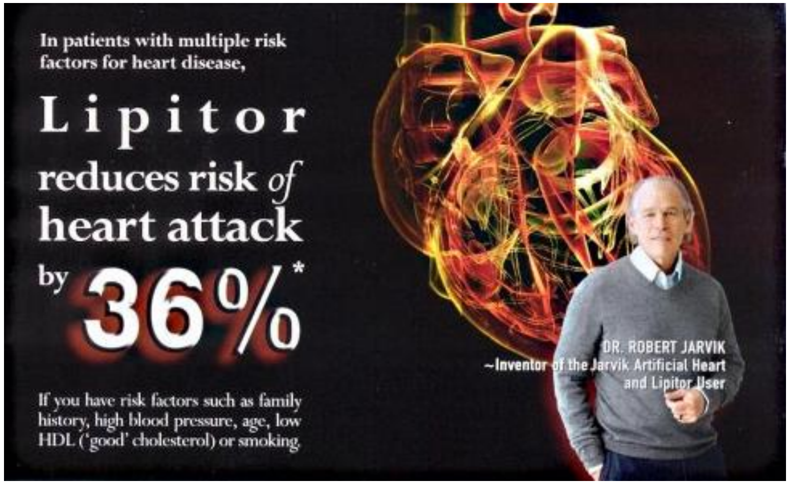
$$NNT = 1 / ARR$$

- NNT, or the number of patients necessary to treat for one patient to benefit, is considered the most useful calculation for assessing treatment benefit

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Relative vs. Absolute Risk Reduction

RRR
(relative risk reduction)



In patients with multiple risk factors for heart disease,

Lipitor reduces risk *of* heart attack by 36%*

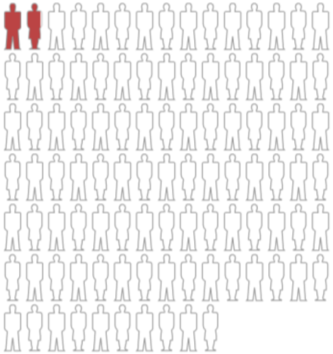
If you have risk factors such as family history, high blood pressure, age, low HDL ('good' cholesterol) or smoking.

DR. ROBERT JARVIK
~Inventor of the Jarvik Artificial Heart and Lipitor User

ARR
(absolute risk reduction)

1.9 vs 3.0 in 100

NNT = 94

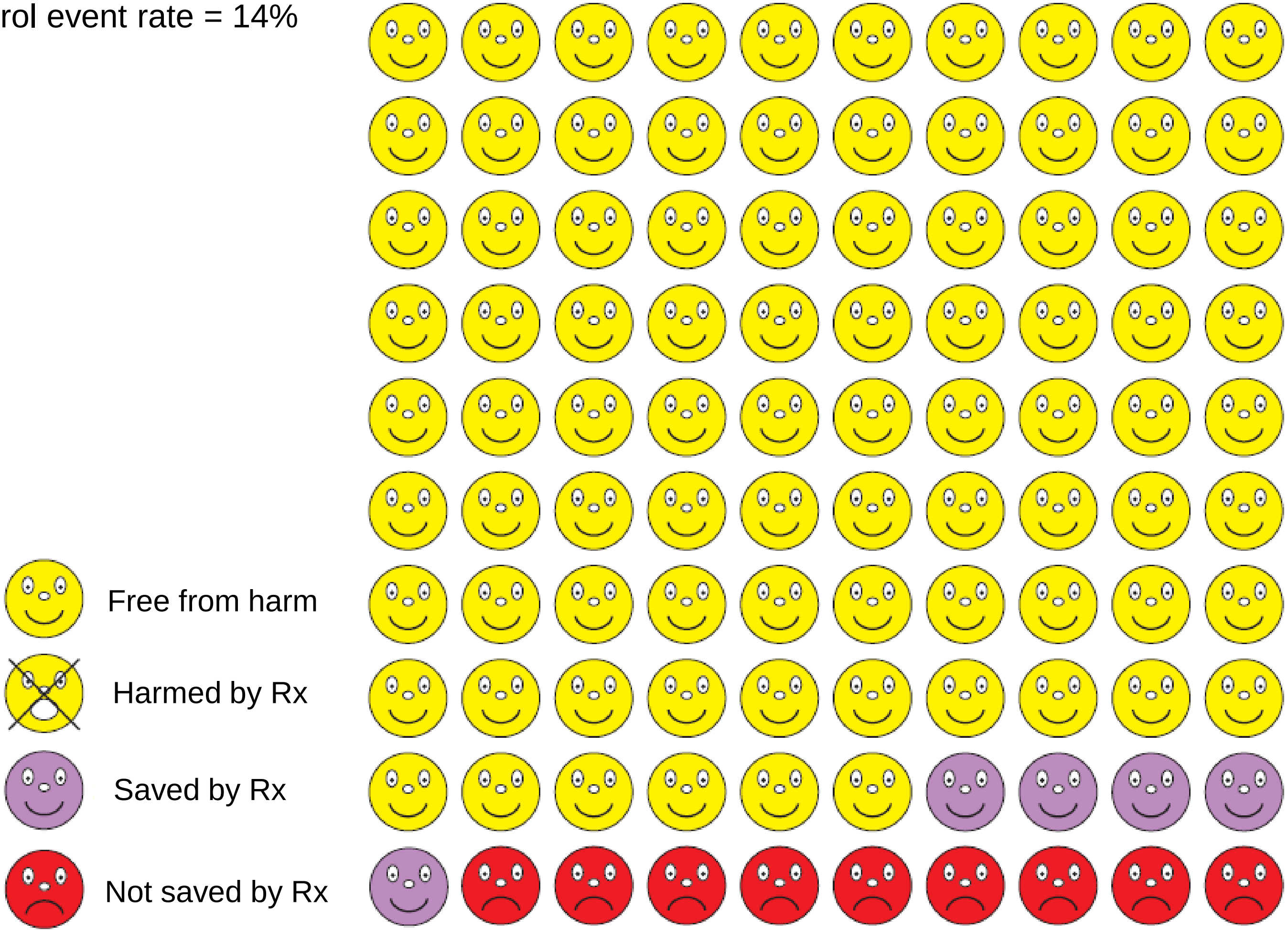


https://acpjc.acponline.org/Content/139/3/issue/ACPJC-2003-139-3-057.htm

Number needed to treat = 20

Pain on days 2-7 in acute otitis media

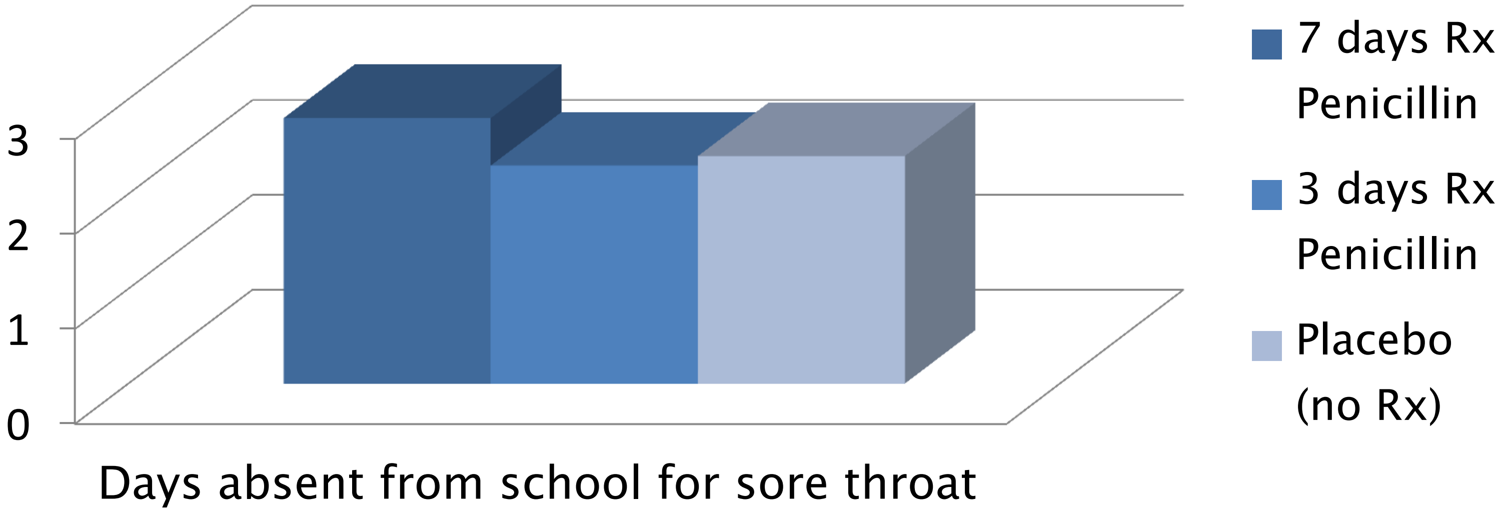
Control event rate = 14%



Portrayal of risks and benefits of treatment with antibiotics for otitis media designed with Visual Rx, a program that calculates numbers needed to treat from pooled results of a meta-analysis and produce graphical display of result

BMJ. 2002 Apr 6;324(7341):827-30. http://www.ncbi.nlm.nih.gov/pubmed/11934777

Penicillin for Sore Throat

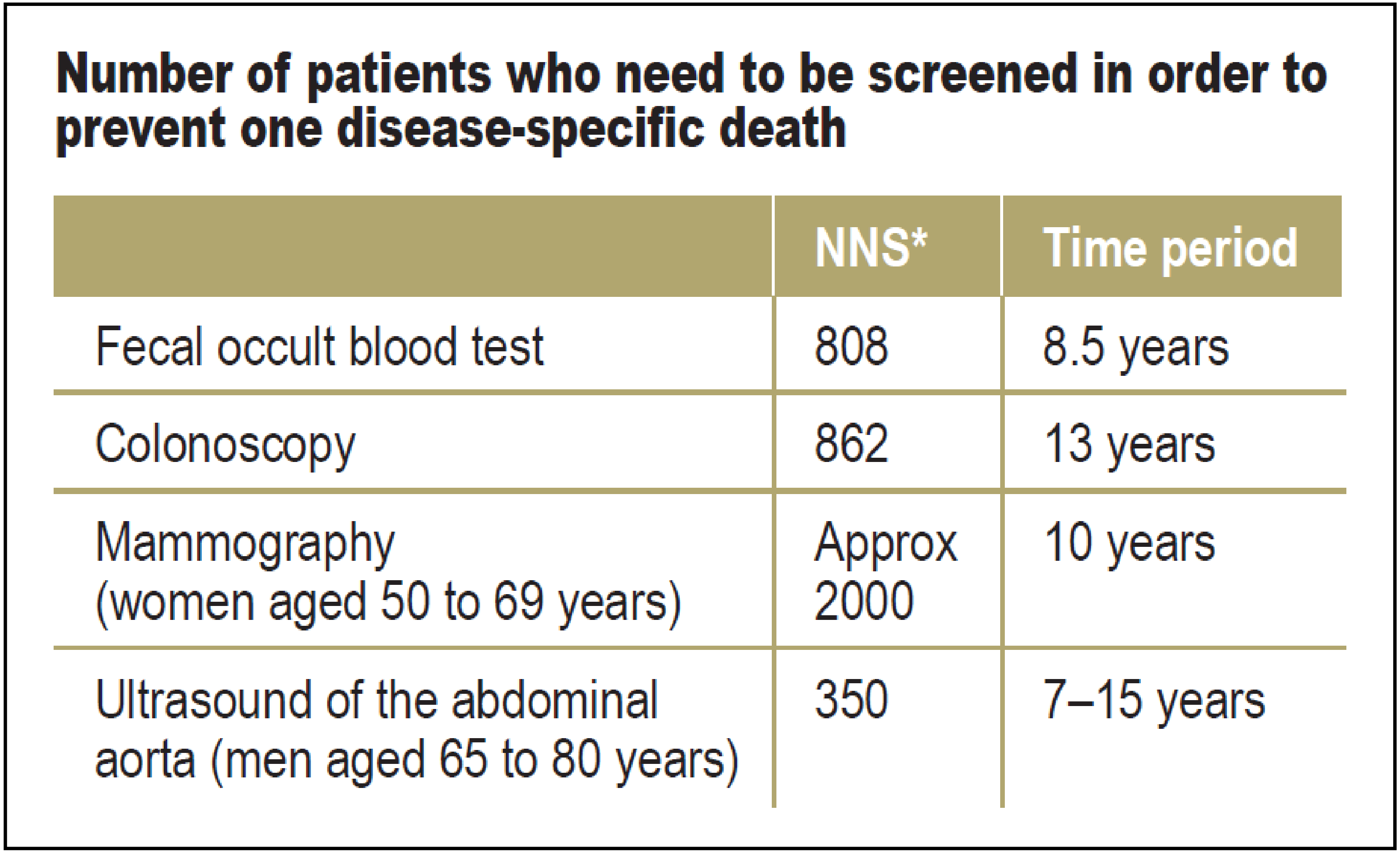


In the total group of children and the children who were positive for group A streptococci, treatment with penicillin for seven days failed to shorten the duration of sore throat, reduce non-attendance at school, or reduce recurrence of sore throat in the following six months.

No rationale exists for treatment with antibiotics in most children with sore throat, irrespective of the presence of streptococci.

Penicillin for acute sore throat in children: randomised, double blind trial
BMJ 12-6-03 http://www.ncbi.nlm.nih.gov/pubmed/14656841

Number Needed to Screen

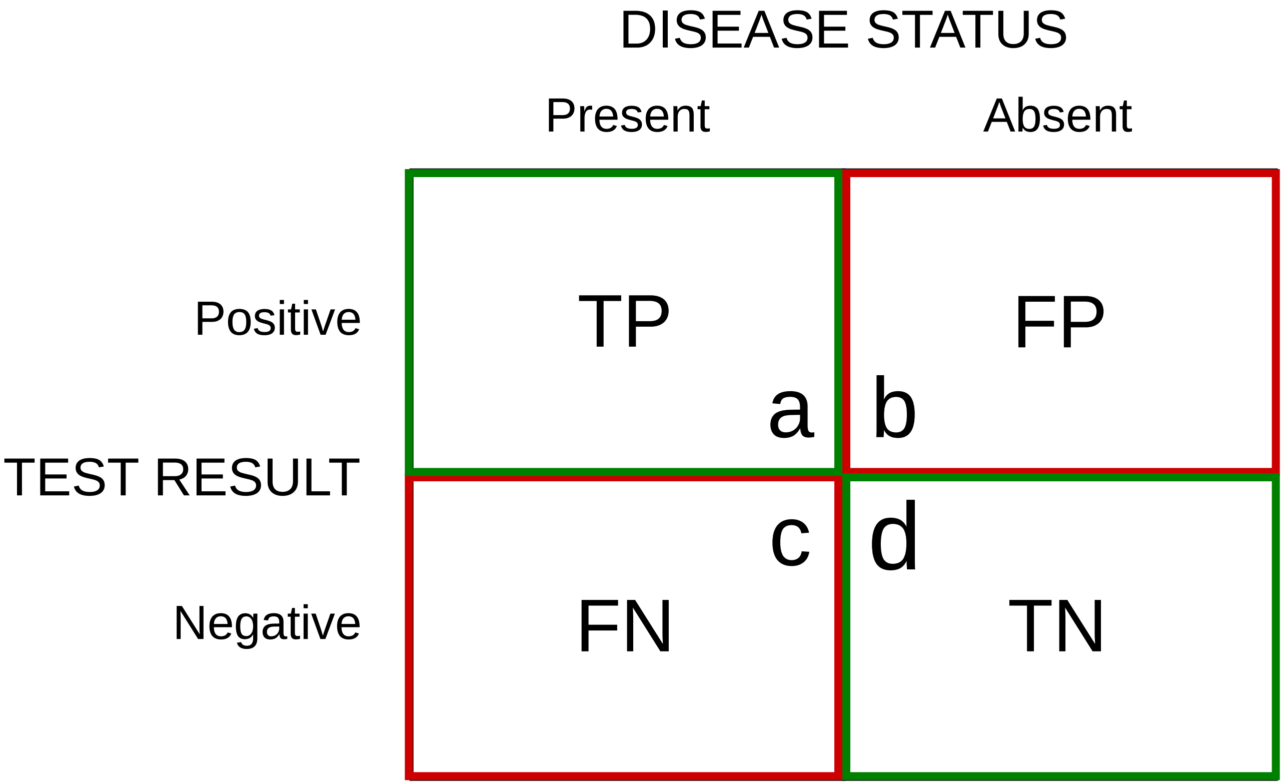


Number of patients who need to be screened in order to prevent one disease-specific death

	NNS*	Time period
Fecal occult blood test	808	8.5 years
Colonoscopy	862	13 years
Mammography (women aged 50 to 69 years)	Approx 2000	10 years
Ultrasound of the abdominal aorta (men aged 65 to 80 years)	350	7–15 years

http://www.ncbi.nlm.nih.gov/pubmed/19946430
http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001877.pub5/pdf

Odds Ratio = $a \ast d \,/\, b \ast c$



		DISEASE STATUS	
		Present	Absent
TEST RESULT	Positive	TP a	FP b
	Negative	c FN	d TN

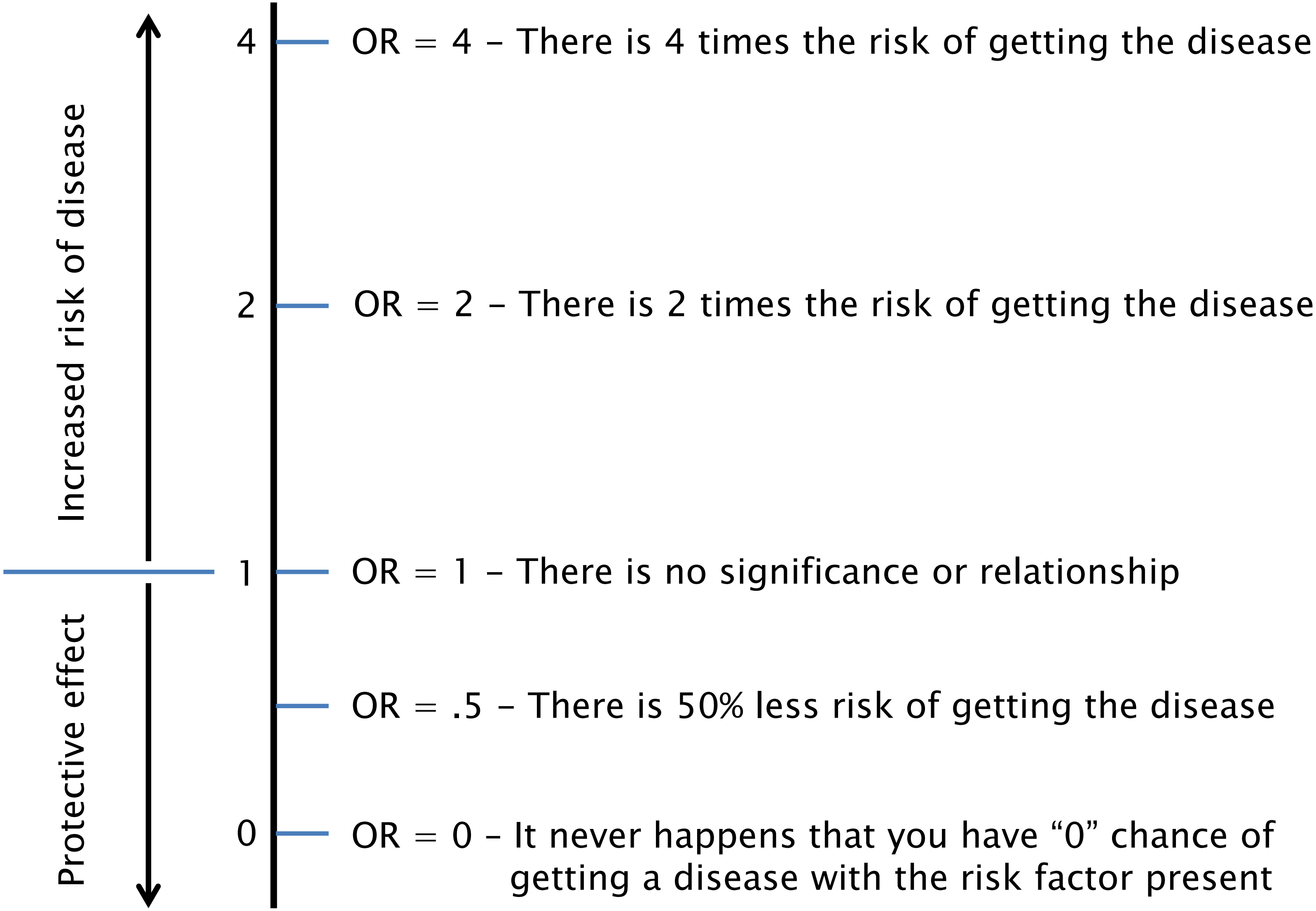
The Odds Ratio, also known as a 2x2 table, is calculated by dividing the True results (TP⁎TN) by the False results (FP⁎FN). It is a measure of the odds of disease in the experimental group vs. the control group.

Odds Ratio

- The odds ratio is the measure of choice in a case-control study
- In a case-control study, the investigator asks both case-patients and a comparison group of persons without disease (“controls”) about their exposures
- Using the information about disease and exposure status, the investigator then calculates an odds ratio to quantify the relationship between exposure and disease

Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Odds Ratio



OR = 4 – There is 4 times the risk of getting the disease

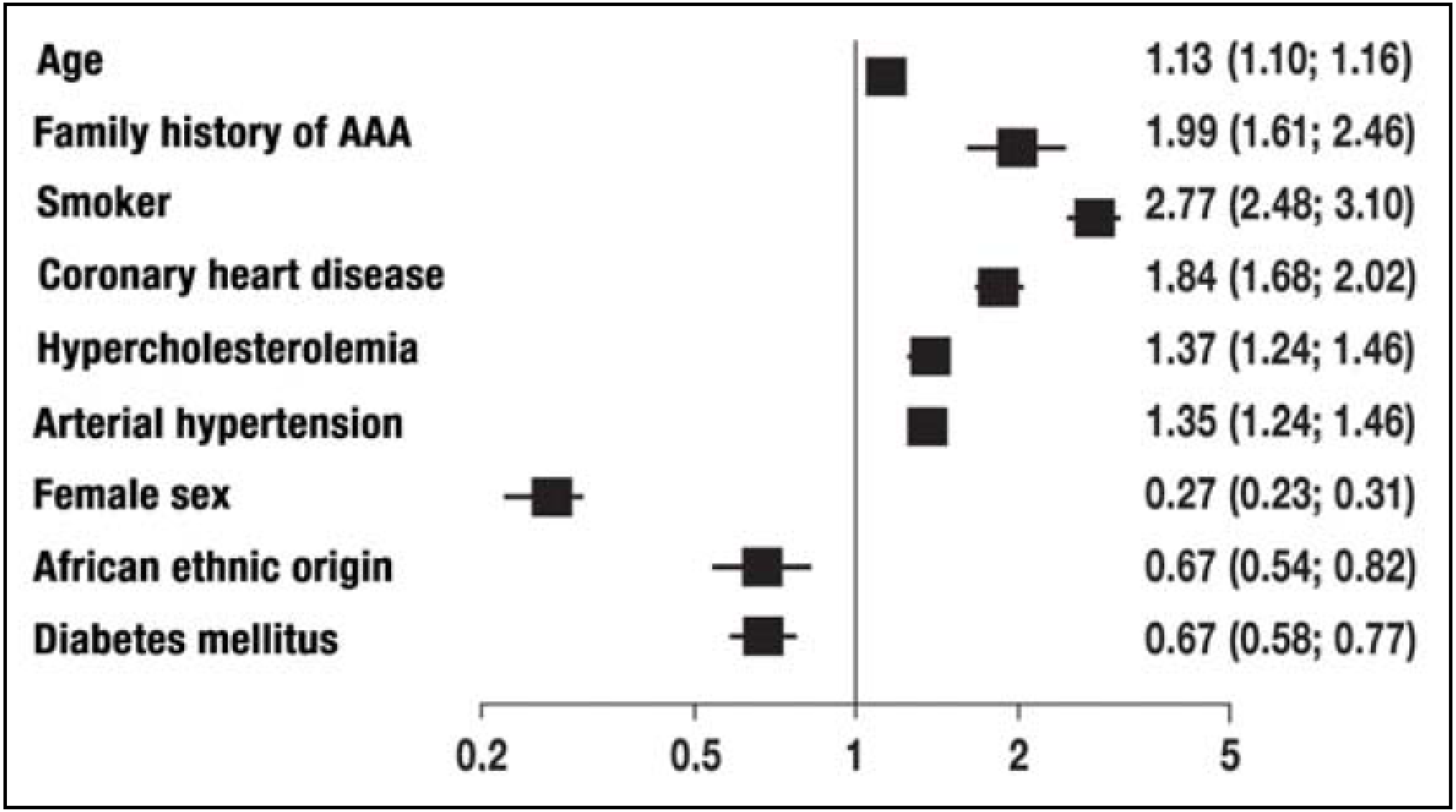
OR = 2 – There is 2 times the risk of getting the disease

OR = 1 – There is no significance or relationship

OR = .5 – There is 50% less risk of getting the disease

OR = 0 – It never happens that you have “0” chance of getting a disease with the risk factor present

Clinical Risk Factors for the Development of an Abdominal Aortic Aneurysm



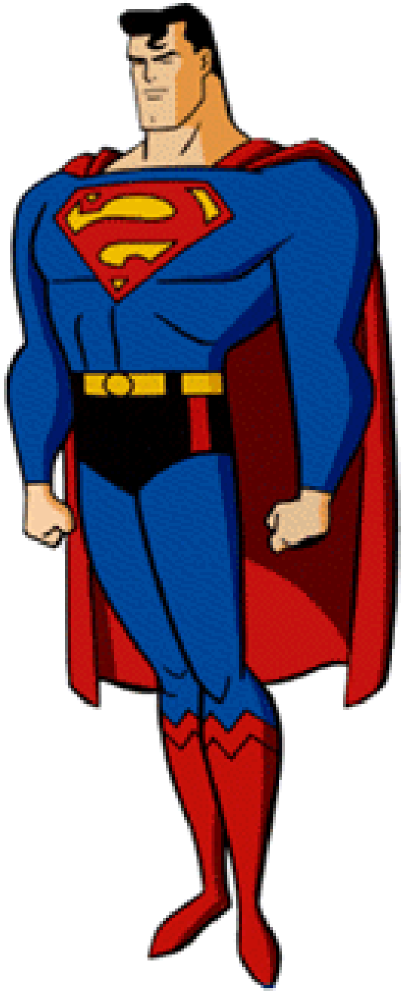
http://www.ncbi.nlm.nih.gov/pubmed/19946430

Confidence Interval

- The confidence interval (CI) represents the range of values consistent with the data from a study and serves as a guide to the variability in a study
- The confidence interval is usually set at 95%, which means that 95% of the true population values will fall within the CI range
- If the confidence interval is narrow, the data is relatively precise and trustworthy
- When the confidence interval is wide, the test results carry less weight

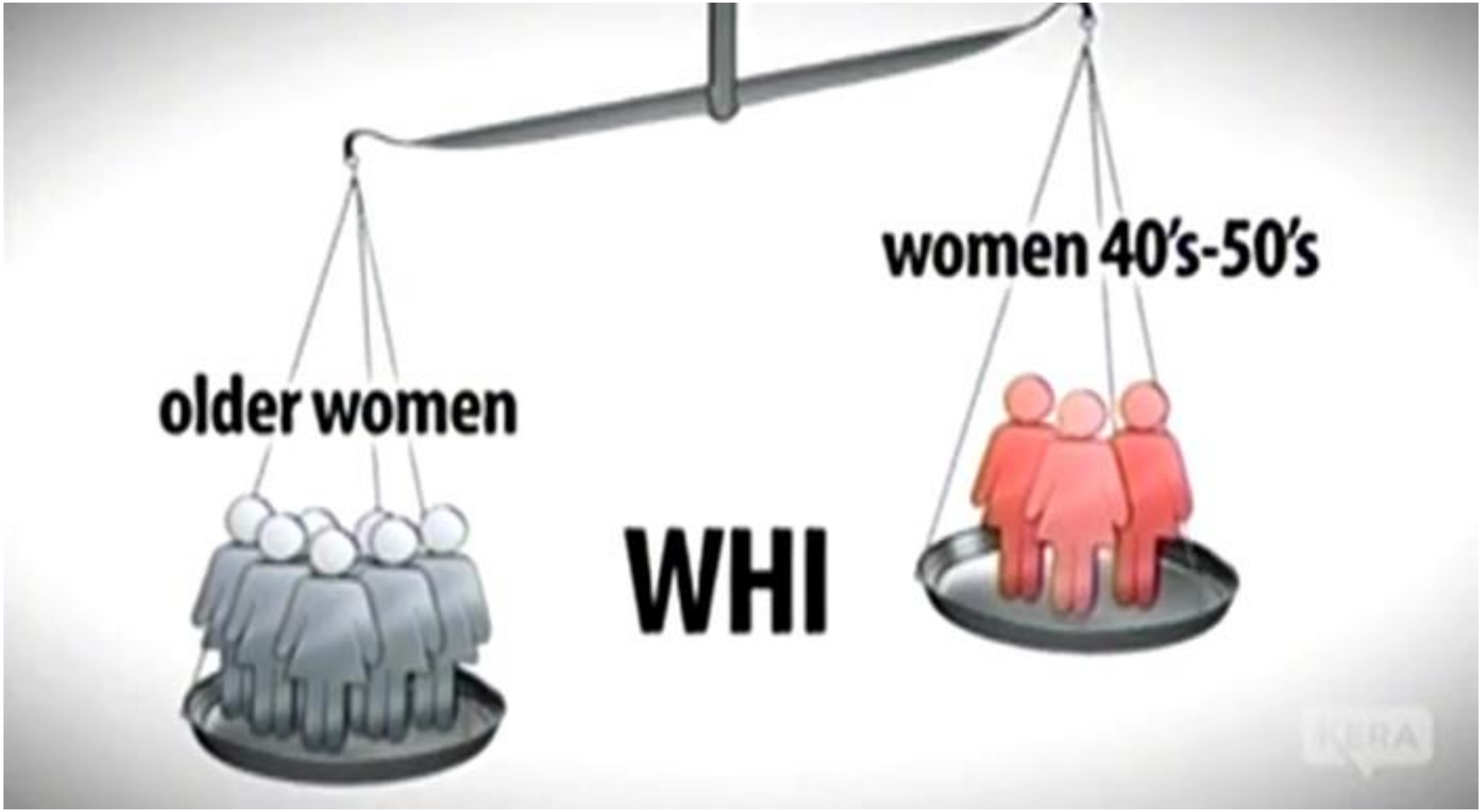
Principles of Epidemiology in Public Health Practice (3rd Edition)
http://www.cdc.gov/training/products/ss1000/ss1000-ol.pdf

Statistical Power



- For a most studies, at least 30 subjects are needed to run tests of statistical significance
- If the sample size is too low, statistical power is compromised and results cannot be trusted or generalized to other patient populations or populations at risk

Women's Health Initiative
(over 160,000 women aged 50–79)



The mortality toll of estrogen avoidance http://www.ncbi.nlm.nih.gov/pubmed/23865654
https://www.youtube.com/watch?v=mewyu54KZVo

Data Mining – The 2002 WHI report

- If women are going to stop taking HRT solely to avoid breast cancer, then, on the basis of the studies to date, they should also stop eating fish, consuming grapefruit, taking antibiotics, using electric blankets, or serving as flight attendants on Scandinavian airlines – **all of which have been reported to have stronger associations with breast cancer than does HRT.**
- Data mining occurs when researchers, having failed to find the statistically significant associations that they had originally hypothesized would exist between a possible risk factor and a disease, go back into their data and rummage around, looking for other factors that might show a statistical link to the dependent variable in question.

Hormone replacement therapy: real concerns and false alarms.
Cancer J. 2009 Mar-Apr;15(2):93-104. http://www.ncbi.nlm.nih.gov/pubmed/19390302

Comorbidity

- Comorbidity is the presence of other diseases or conditions that may affect the individual
- These additional health problems will often confound treatment results and need to be adjusted for in research studies
- A recent study found that daily back pain doubled the risk of coronary heart disease and all cause mortality
- The effect remained significant even after being adjusted for physical activity and cardiovascular risk factors (BMI, smoking, HTN, hypercholesterolemia, diabetes, etc.)

Spine 2007 Aug 15;32(18):2012-8. http://www.ncbi.nlm.nih.gov/pubmed/17700450

Back Pain and Mortality

Back Pain	Age adjusted	CV risk factors adjusted	Physical activity adjusted
Infrequent (<1/mo)	1.00	1.00	1.00
Frequent (1/mo to 1/day)	OR=1.27 CI (0.70–2.32)	OR=1.21 CI (0.66–2.22)	OR=1.24 CI (0.68–2.26)
Daily (1/day or more)	OR=2.03 CI (1.14–3.60)	OR=1.85 CI (1.00–3.43)	OR=1.93 CI (1.08–3.43)

Spine 2007 Aug 15;32(18):2012-8. http://www.ncbi.nlm.nih.gov/pubmed/17700450

Statistical Illiteracy

- Is common to patients, physicians, and politicians
- Is created by nontransparent framing of information that is sometimes an unintentional result of lack of understanding
- Can also be a result of intentional efforts to manipulate or persuade people
- Can have serious consequences for health

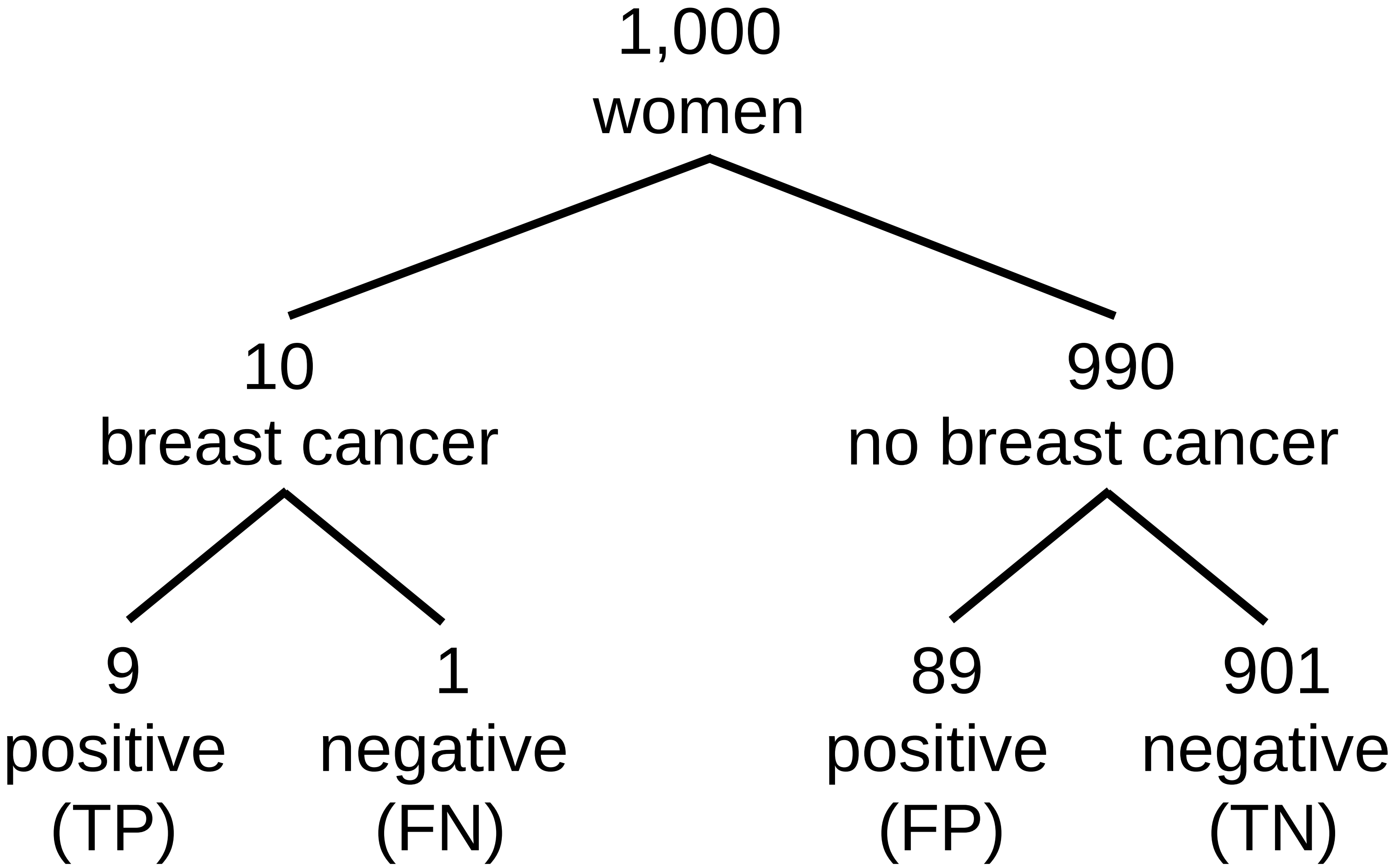
http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Remedy for Statistical Illiteracy

Transparent Information	Non-Transparent Information
Natural Frequencies	Conditional Probabilities (sensitivity, specificity)
Mortality Rates	Survival Rates
Absolute Risks	Relative Risks

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Natural Frequencies



1,000
women

10
breast cancer

990
no breast cancer

9
positive
(TP)

1
negative
(FN)

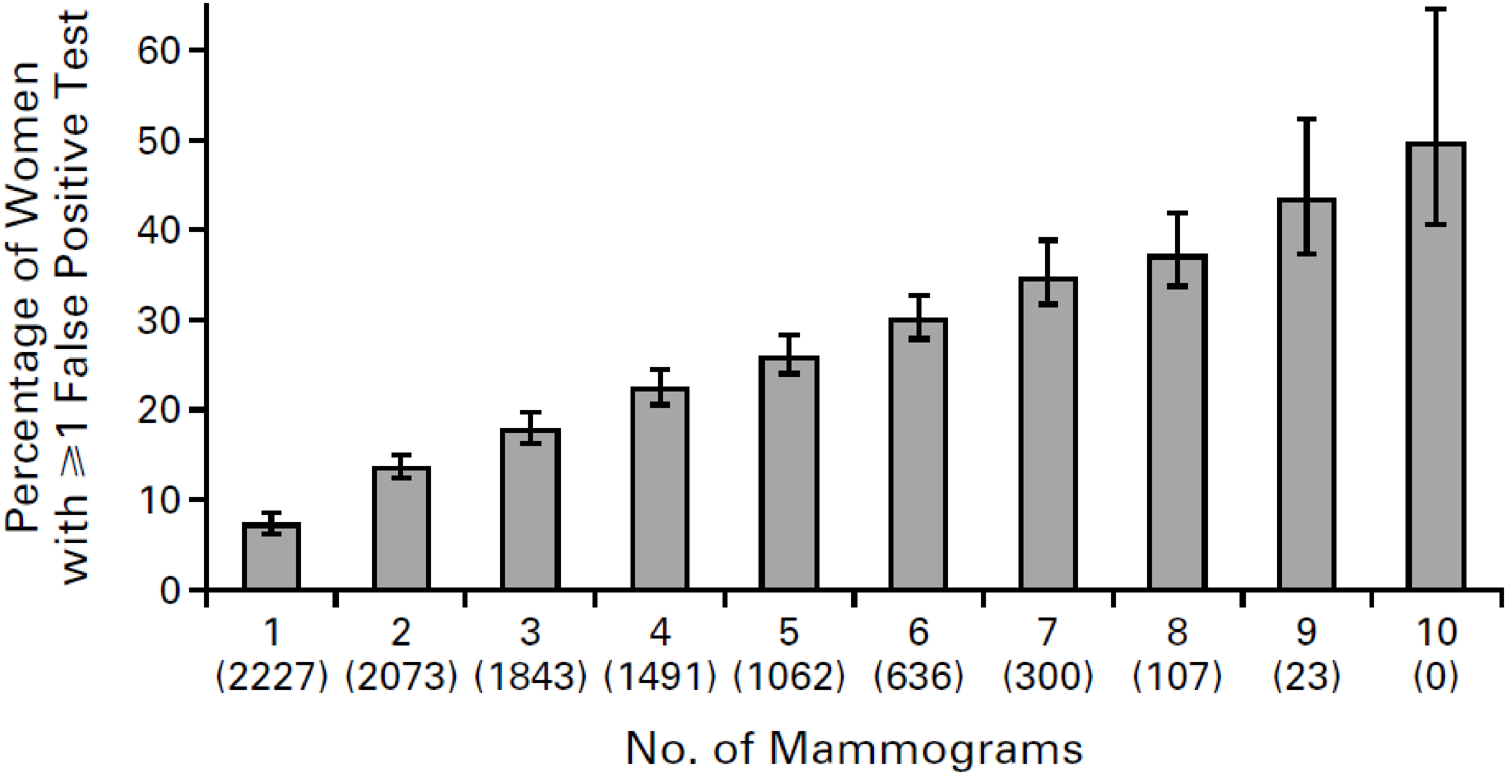
89
positive
(FP)

901
negative
(TN)

Conditional Probabilities

		DISEASE STATUS		
		Present	Absent	
TEST RESULT	Positive	TP = 9	FP = 89	PPV TP / (TP+FP) = 9.2%
	Negative	FN = 1	TN = 901	NPV TN / (TN+FN) = 99.9%
		Sensitivity TP / (TP+FN) = 90%	Specificity TN / (TN+FP) = 91%	

Screening Mammography



The risk of a false positive test increased with the number of breast-cancer screening tests, so that by the time a woman had undergone 10 tests, the estimated cumulative risk of at least one false positive mammogram was about 50 percent

N Engl J Med. 1998 Apr 16;338(16):1089-96. www.ncbi.nlm.nih.gov/pubmed/9545356

Multiple-Choice Question

1. Regular mammography in women who are well:
   a) prevents the risk of contracting breast cancer
   b) reduces the risk of contracting breast cancer
   c) does not have any influence on the risk of contracting breast cancer

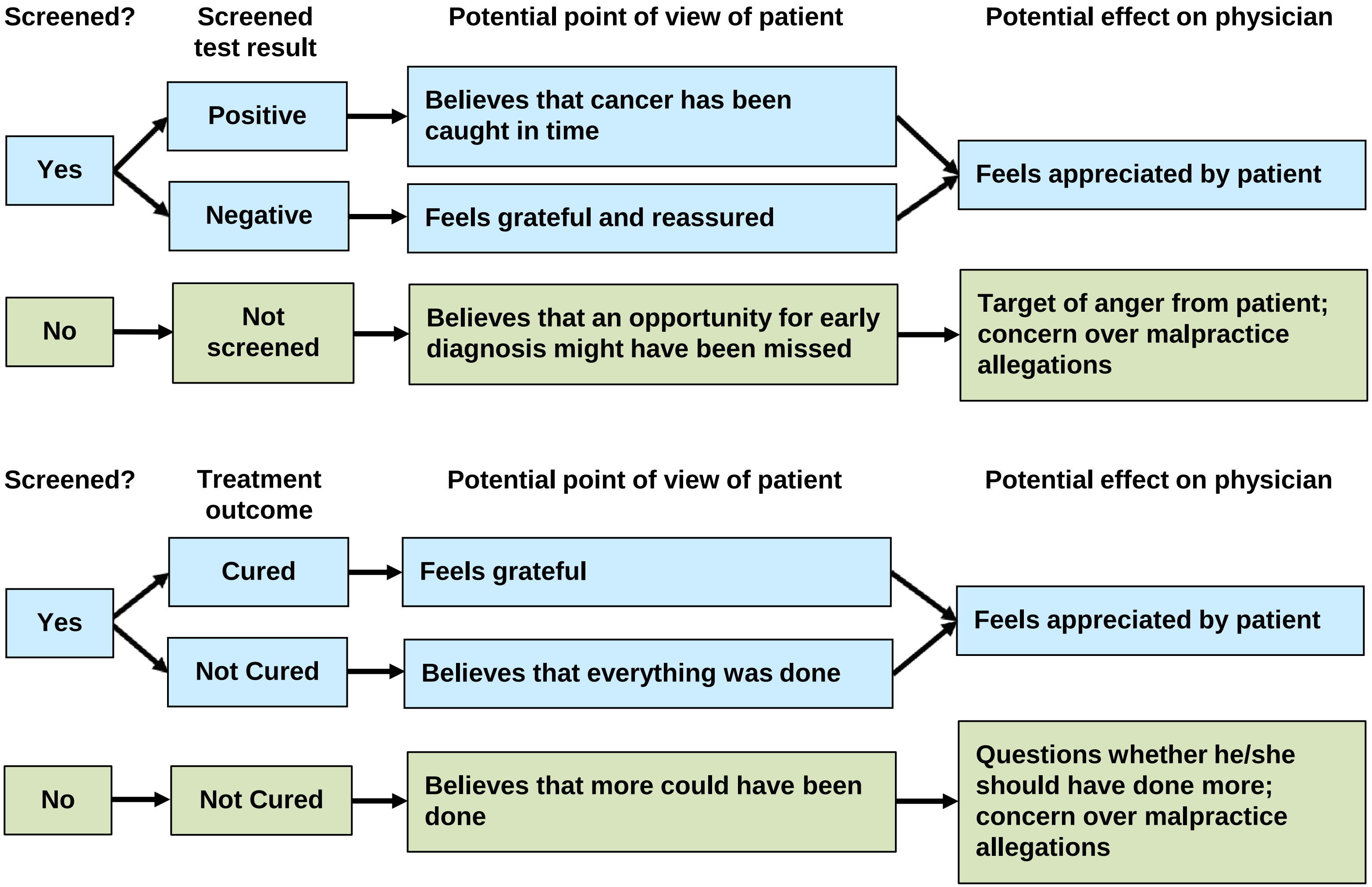
http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Screening Tests

- Screening is testing for hidden disease in people without symptoms
- The goal of screening is early detection, it has no effect on whether you get the disease
- Due to over-diagnosis and lead-time bias, changes in 5 year survival rates have no reliable relationship to changes in mortality

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Over-Diagnosis Bias



Screened?

Screened test result

Potential point of view of patient

Potential effect on physician

Yes

Positive

Believes that cancer has been caught in time

Negative

Feels grateful and reassured

Feels appreciated by patient

No

Not screened

Believes that an opportunity for early diagnosis might have been missed

Target of anger from patient; concern over malpractice allegations

Screened?

Treatment outcome

Potential point of view of patient

Potential effect on physician

Yes

Cured

Feels grateful

Not Cured

Believes that everything was done

Feels appreciated by patient

No

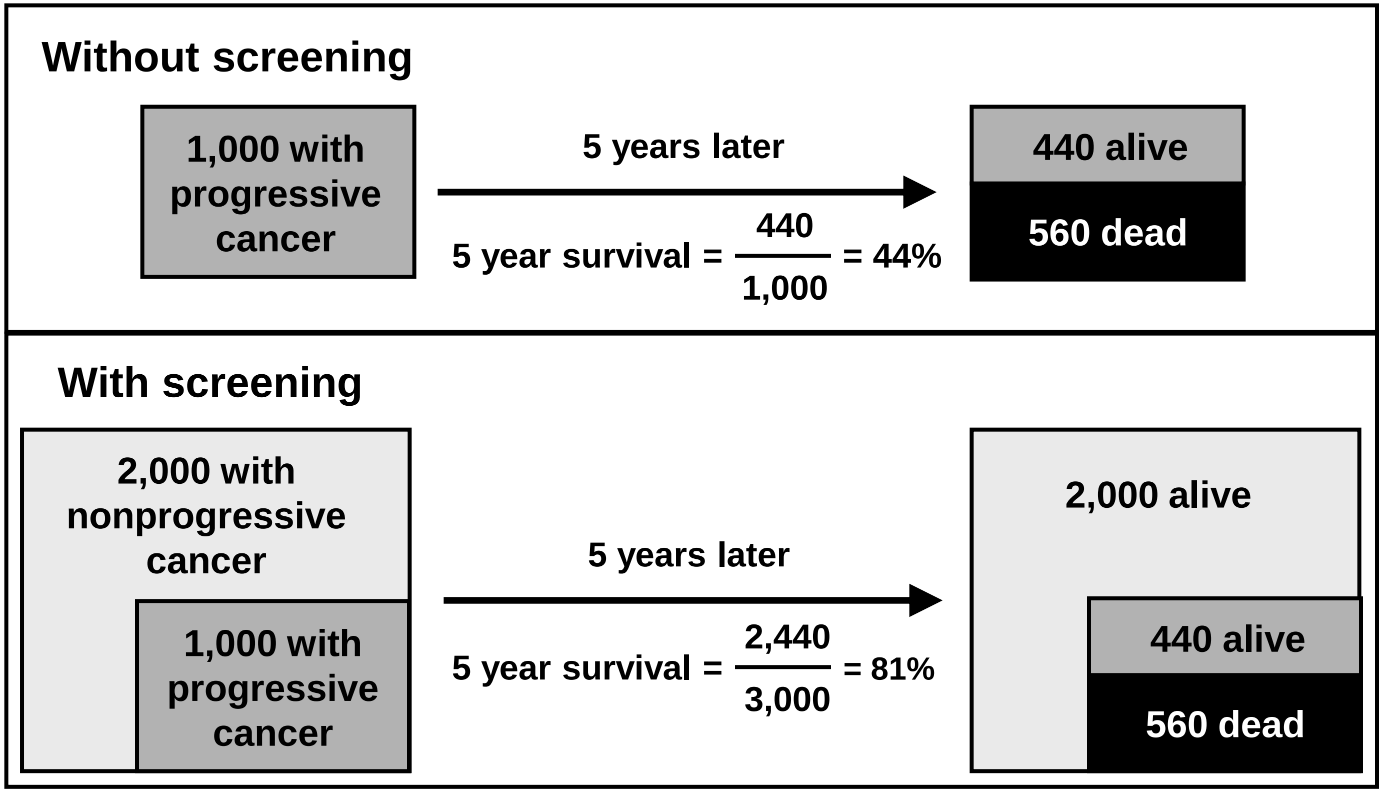
Not Cured

Believes that more could have been done

Questions whether he/she should have done more; concern over malpractice allegations

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4322920/

Over-Diagnosis Bias



Without screening

1,000 with progressive cancer

5 years later

5 year survival = $\frac{440}{1{,}000}$ = 44%

440 alive

560 dead

With screening

2,000 with nonprogressive cancer

1,000 with progressive cancer

5 years later

5 year survival = $\frac{2{,}440}{3{,}000}$ = 81%

2,000 alive

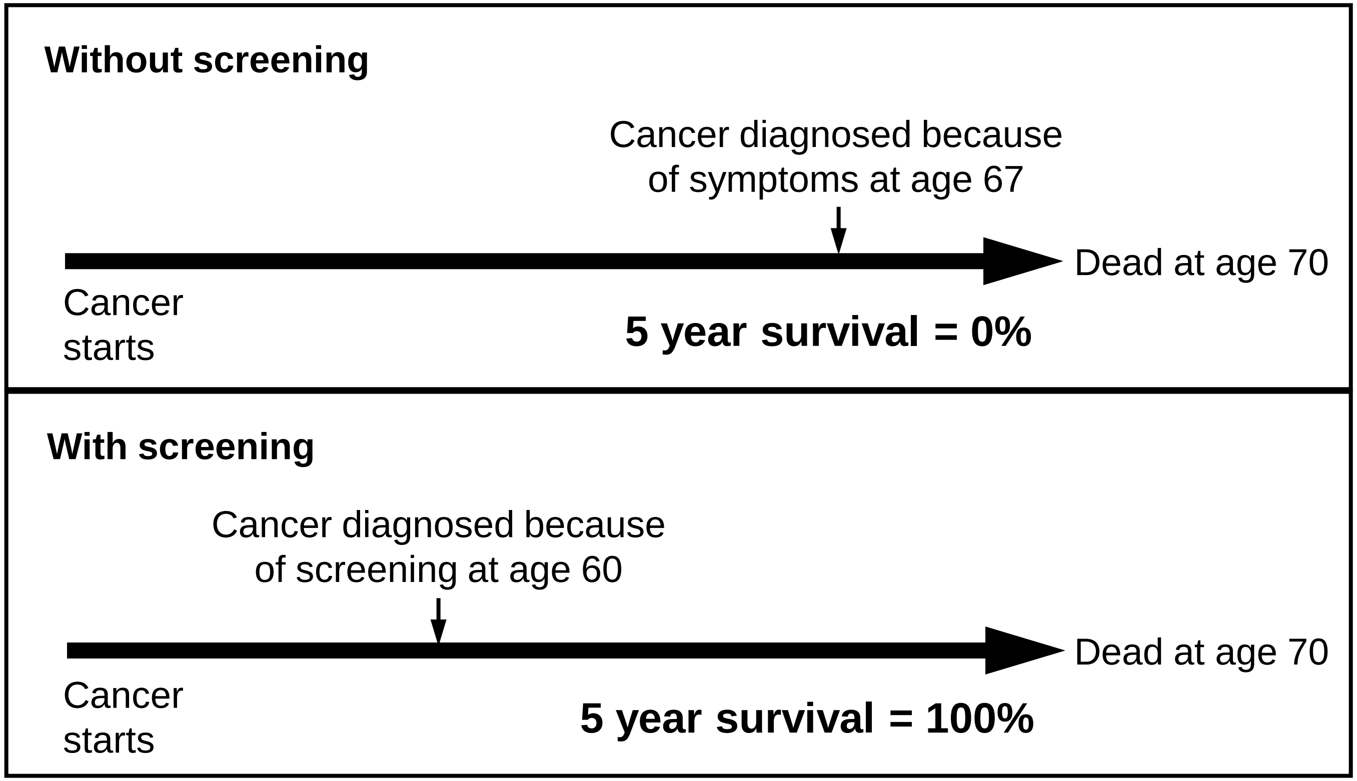
440 alive

560 dead

Even if the number of people who die is not changed by screening, and thus no life is saved or prolonged, screening detected nonprogressive cancers can inflate the 5 year survival rates, causing such statistics to be misleading.

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Lead-Time Bias



Without screening

Cancer diagnosed because
of symptoms at age 67

Dead at age 70

Cancer
starts

5 year survival = 0%

With screening

Cancer diagnosed because
of screening at age 60

Dead at age 70

Cancer
starts

5 year survival = 100%

Even if the time of death is not changed by screening, and thus no life is saved or prolonged, advancing the time of diagnosis in this way can result in increased 5 year survival rates, causing such statistics to be misleading.

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Survival vs. Mortality Rates

Survival Rate

- *“The 5 year survival rate for people diagnosed with prostate cancer is 98% in the USA vs. 71% in Britain.”*

Mortality Rate

- *“There are 26 prostate cancer deaths per 100,000 American men vs. 27 per 100,000 men in Britain.”*

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Absolute vs. Relative Risk

Relative risk:

- If you have this test every 2 years, it will reduce your chance of dying from this cancer by around one third over the next 10 years

Absolute risk:

- If you have this test every 2 years, it will reduce your chance of dying from this cancer from around 3 in 1,000 to around 2 in 1,000 over the next 10 years

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Absolute vs. Relative Risk

Relative risk:

- Drug X reduced the risk of dying from heart attack by 50%

Absolute risk:

- Drug X reduced the risk of dying from heart attack from 200 to 100 in 10,000 patients, or
- Drug X reduced the risk of dying from heart attack from 2 to 1 in 10,000 patients

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Mismatched Framing of Information

- An analysis of three major medical journals, BMJ, JAMA, and The Lancet from 2004 to 2006 revealed that when both benefits and harms of therapeutic interventions were reported, 1 in 3 studies used mismatched framing
- Benefits are reported using relative risk
- Harms are reported using absolute risk

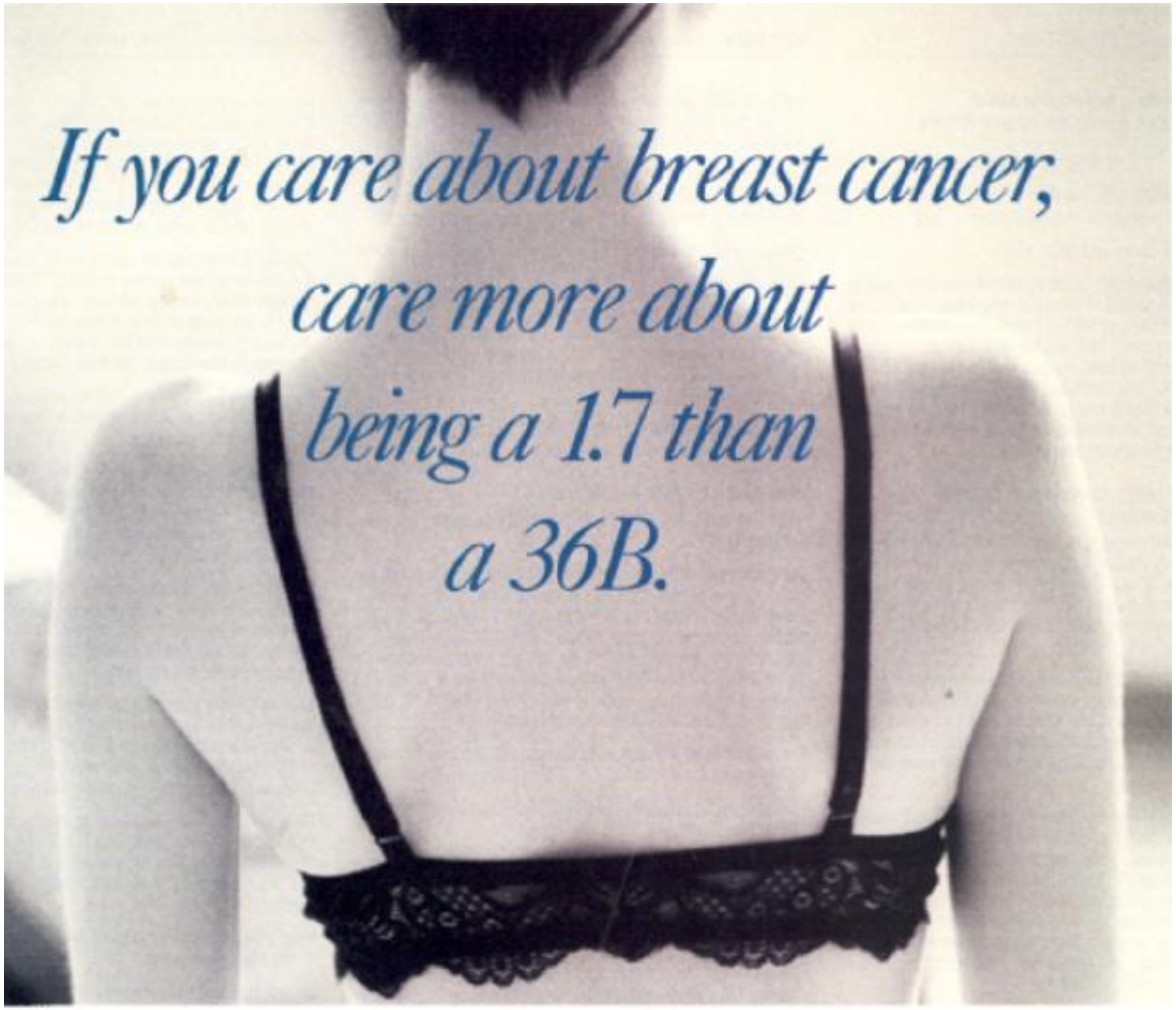
http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Mismatched Framing of Information

For example:

- Benefits – relative risk (large numbers)
  - *"Sigmoidoscopy screening reduced the risk of death by 59% for cancers within reach of the sigmoidoscope"*
- Harms – absolute risk (small numbers)
  - *"Perforations are reported to occur in approximately 1 of 1,000-10,000 rigid sigmoidoscopic examinations"*

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf



Know your breast cancer risk assessment number.
Know that NOLVADEX (tamoxifen citrate) could reduce your
chances of getting breast cancer if you are at high risk.

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

13,000 women at high risk of getting breast cancer were given either NOLVADEX or a sugar pill for 5 years. Here's what happened:

What difference did NOLVADEX make?	Women given a sugar pill	Women given NOLVADEX
Fewer women got invasive breast cancer (16 in 1,000 fewer due to drug)	3.3% 33 in 1,000	1.7% 17 in 1,000
No difference in dying from breast cancer	0.09% 0.9 in 1,000	0.09% 0.9 in 1,000
More women had a blood clot in their leg or lungs (additional 5 in 1,000 due to drug)	0.5% 5 in 1,000	1.0% 10 in 1,000
More women got invasive uterine cancer (additional 6 in 1,000 due to drug)	0.5% 5 in 1,000	1.1% 11 in 1,000
More women had cataracts needing surgery (additional 8 in 1,000 due to drug)	1.5% 15 in 1,000	2.3% 23 in 1,000

http://www.psychologicalscience.org/journals/pspi/pspi_8_2_article.pdf

Observational Studies

- The widely held view that randomised controlled trials are the "gold standard" for evaluation has led to the denigration of non-experimental methods, to the extent that research funding bodies and journal editors automatically reject them
- There are some people who believe that any and every intervention can be subjected to a randomised trial, and that those who challenge this have simply not made sufficient effort and are methodologically incompetent

BMJ. 1996 May 11;312(7040):1215-8. http://www.ncbi.nlm.nih.gov/pubmed/8634569

Observational Studies

- For too long a false conflict has been created between those who advocate randomised trials in all situations and those who believe observational data provide sufficient evidence.
- Neither position is helpful. There is no such thing as a perfect method; each method has its strengths and weaknesses.

BMJ. 1996 May 11;312(7040):1215-8. http://www.ncbi.nlm.nih.gov/pubmed/8634569

Efficacy vs. Effectiveness

- The research question remains the same for both: *"What works? For whom?"* and *"When?"*
- The *efficacy* of an intervention is what works under "ideal and controlled" conditions, whereas the *effectiveness* of an intervention is what works under "everyday and real-life" conditions
- It is clear that for the question of efficacy, the RCT is unarguably the best design; however, given that randomization and controlled conditions play no part in everyday clinical practice, evidence for effectiveness cannot be accrued through use of the same design

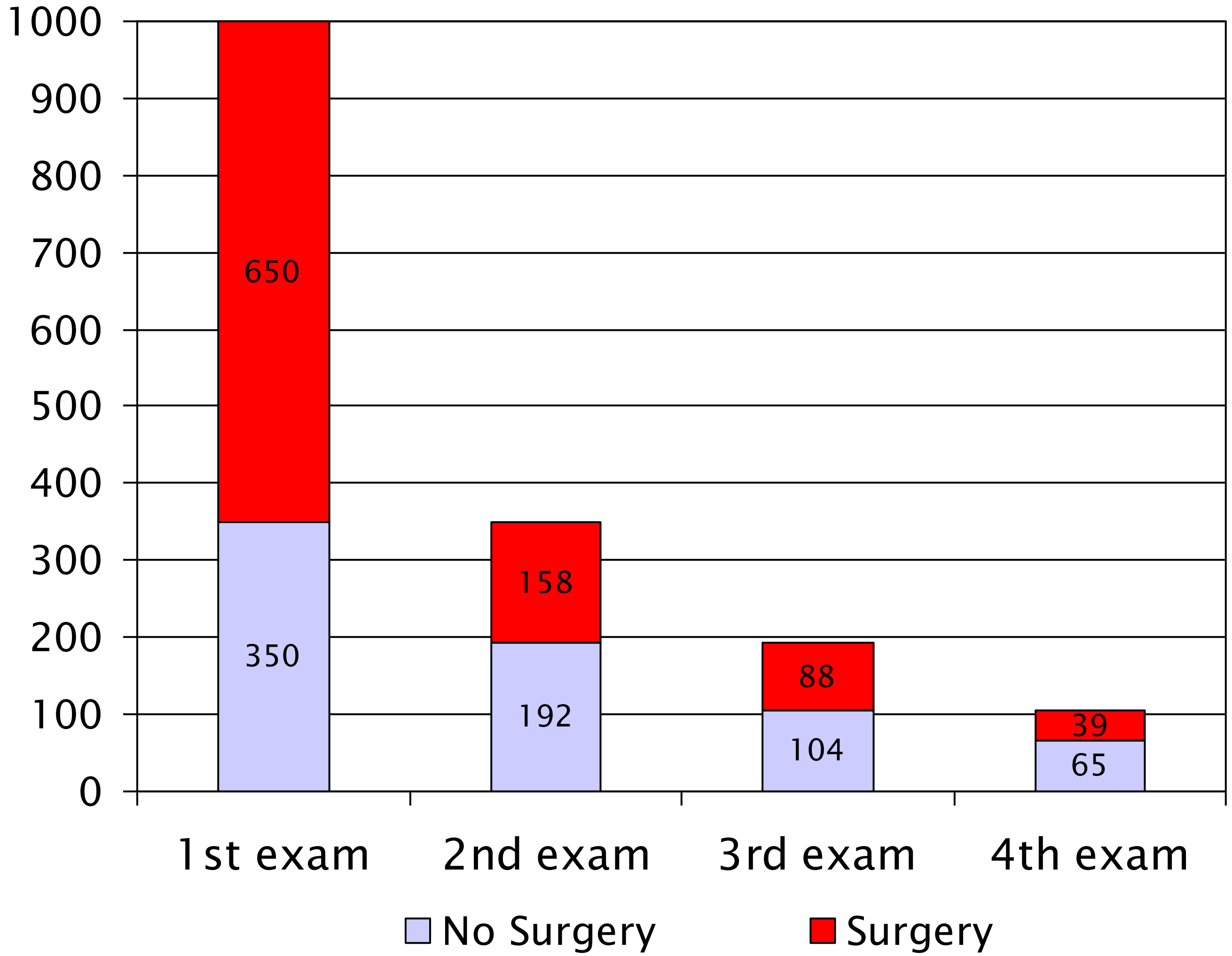
J Manipulative Physiol Ther. 2001 Jun;24(5):362-6. www.ncbi.nlm.nih.gov/pubmed/11416828

Tonsillectomy Study

- A survey of 1,000 New York City school children found that 65% had undergone a tonsillectomy
- The remaining children were examined and 45% were selected for tonsillectomy
- Those rejected were examined by another group of physicians and 46% were selected for surgery
- When the process was repeated again with a different group of physicians, only 65 students from the original 1,000 students were left
- At that point, the study was halted for lack of physicians

http://content.healthaffairs.org/cgi/reprint/3/2/74.pdf

Tonsillectomy Study



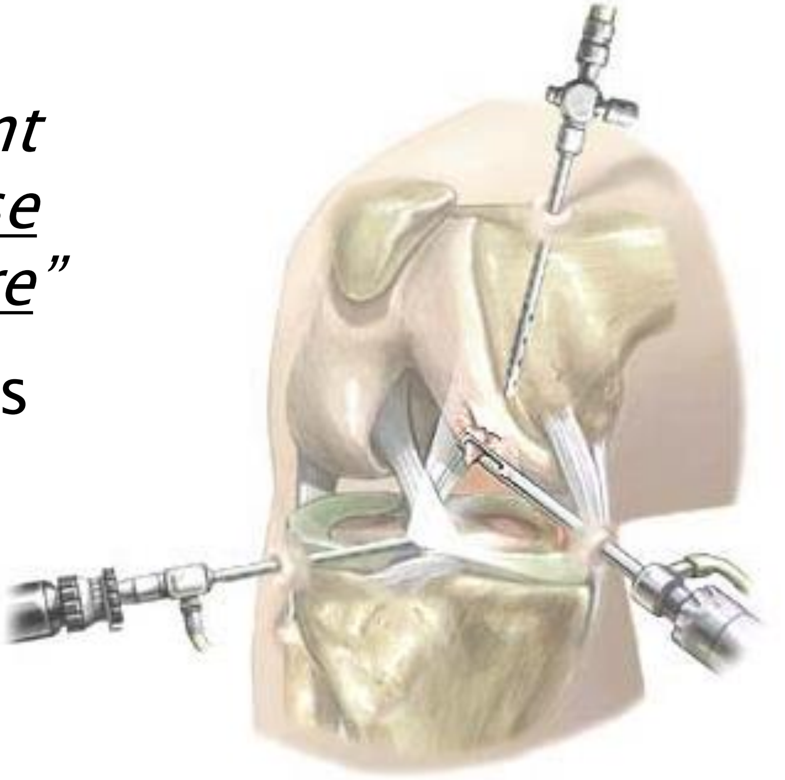
Placebo Effect of Surgery

- In the 1950’s there was a surgical procedure for angina: ligation of the internal mammary artery
- It was theorized that this would shunt more blood to the coronary arteries
- A Reader’s Digest article, *“New Surgery for Ailing Hearts”* sent many patients to surgeons asking for the operation
- When the artery ligation procedure was tested against sham surgery, both groups had the same reduction in angina pain

http://books.google.com/books?id=SiJ2UB3dv9UC&pg=PA448
http://www.nap.edu/openbook.php?record_id=607&page=17

Arthroscopic Surgery of the Knee

- *"… the outcomes after arthroscopic lavage or arthroscopic débridement were no better than those after a placebo procedure"*
- 650,000 such procedures are performed each year at a cost of roughly $5,000 each
- = $3.25 Billion per year for surgery that was no better than placebo



A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee
NEJM 7-11-02 http://www.ncbi.nlm.nih.gov/pubmed/12110735

“Do a double-blind test. Give the new drug to rich patients and a placebo to the poor. No sense getting their hopes up. They couldn’t afford it even if it works.”